

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2001, 23:05:42 ; Search time 7860.64 Seconds  
(without alignments)  
1923.232 Million cell updates/sec

Title: US-09-477-392-1  
Perfect score: 2954  
Sequence: 1 cgcgttgcacgctctc.....aaaaaaaaaaaaaaaaaaaaa 2954

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues  
tal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pl1:\*  
9: gb\_pr2:\*  
10: gb\_pr3:\*  
11: gb\_ro:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: em\_fun:\*  
15: em\_hum1:\*  
16: em\_hum2:\*  
17: em\_in:\*  
18: em\_cm:\*  
19: em\_or:\*  
20: em\_ov:\*  
21: em\_pat:\*  
22: em\_ph:\*  
23: em\_pl:\*  
24: em\_ro:\*  
25: em\_sts:\*  
26: em\_sy:\*  
27: em\_un:\*  
28: em\_vl:\*  
29: gb\_htg1:\*  
30: gb\_htg2:\*  
31: gb\_in1:\*  
32: gb\_in2:\*  
33: em\_da1:\*  
34: em\_da2:\*  
35: em\_hum3:\*  
36: em\_hum4:\*  
37: gb\_pr4:\*  
38: gb\_htg3:\*  
39: gb\_htg4:\*  
40: gb\_htg5:\*  
41: gb\_htg6:\*  
42: gb\_htg7:\*  
43: em\_htg1:\*

44: em\_htg2:\*  
45: em\_htg3:\*  
46: em\_hum5:\*  
47: gb\_pl3:\*  
48: gb\_pr5:\*  
49: gb\_htg8:\*  
50: gb\_htg9:\*  
51: gb\_htg10:\*  
52: gb\_htg11:\*  
53: gb\_htg12:\*  
54: gb\_htg13:\*  
55: gb\_in3:\*  
56: gb\_in3:\*  
57: gb\_htg15:\*  
58: gb\_htg16:\*  
59: gb\_htg17:\*  
60: em\_htg4:\*  
61: em\_htg5:\*  
62: em\_htg6:\*  
63: em\_htg7:\*  
64: em\_hum6:\*  
65: gb\_htg18:\*  
66: gb\_htg19:\*  
67: gb\_htg20:\*  
68: gb\_htg21:\*  
69: gb\_htg22:\*  
70: gb\_htg23:\*  
71: gb\_vil1:\*  
72: gb\_vil2:\*  
73: gb\_ba3:\*  
74: em\_htg8:\*  
75: em\_htg9:\*  
76: em\_htg10:\*  
77: gb\_pr6:\*  
78: gb\_pr7:\*  
79: gb\_sts1:\*  
80: gb\_sts2:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_htg0:\*  
84: gb\_htg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2954	100.0	2954	37 AF201303	AF201303 Homo sapi
2	2729	92.4	132150	9 AC005586	AC005586 Homo sapi
3	2719.4	92.1	2878	37 AK025356	AK025356 Homo sapi
4	2654.6	89.9	2878	77 HSA245553	AJ245553 Homo sapi
5	1256.4	42.5	179225	58 AC073111	AC073111 Homo sapi
6	1194.8	40.4	201458	40 AC015887	AC015887 Mus muscu
7	719	24.3	759	78 HSM802596	AL162065 Homo sapi
8	315.8	10.7	6045	8 AB002324	AB002324 Human mRN
9	315.8	10.7	185306	39 AC013570	AC013570 Homo sapi
10	315.4	10.7	239566	77 HSAJ03147	AJ03147 Homo sapi
11	315.2	10.7	2382	37 AK023017	AK023017 Homo sapi
12	310.4	10.5	130067	9 AC007228	AC007228 Homo sapi
13	304.6	10.3	2094	11 MM2EP29	X55126 M.musculus
14	299.8	10.1	183525	39 AC012313	AC012313 Homo sapi
15	283.6	9.6	230534	65 AC079575	AC079575 Mus muscu
16	282.8	9.6	128361	8 AC004877	AC004877 Homo sapi
17	281	9.5	203100	55 AC060772	AC060772 Mus muscu
18	279.4	9.5	295	78 HSG5	X82192 H.sapiens E
19	278.4	9.4	2311	78 HS2NE74A3	AF072567 Homo sapi
20	278.4	9.4	2416	78 HS2NE741	X71623 H.sapiens Z
21	278.4	9.4	2416	79 G27154	G27154 SHGC-31580

22	278.2	9.4	3759	78	HSZNF742	X92715 H.sapiens m
23	278.4	9.4	183617	8	AC007731	AC007731 Homo sapi
24	278.4	9.4	192592	8	AC005500	AC005500 Homo sapi
25	277.8	9.4	4582	11	MUSPLM4X	M98502 Mus musculus
26	277.6	9.4	2587	10	AF055077	AF055077 Homo sapi
27	277.6	9.4	2678	85	HUMMZFI	M58297 Human zinc
28	277.6	9.4	2905	10	AF055078	AF055078 Homo sapi
29	277.6	9.4	12456	37	AF161886	AF161886 Homo sapi
30	277.6	9.4	46616	41	AC016630	AC016630 Homo sapi
31	277.6	9.4	44210	41	AC016639	AC016639 Homo sapi
32	275.6	9.3	2869	11	AB007407	AB007407 Mus muscu
33	274.2	9.3	113247	65	AC079506	AC079506 Mus muscu
34	273.8	9.3	1554	37	AF269249	AF269249 Homo sapi
35	273.8	9.3	130067	9	AC007228	AC007228 Homo sapi
36	270.2	9.1	210923	39	AC013273	AC013273 Homo sapi
37	263.8	8.9	138278	84	HSJ99615	AL109966 Homo sapi
38	260.6	8.8	166803	55	AC048332	AL048382 Homo sapi
39	260.6	8.8	194486	41	AC016771	AC016771 Homo sapi
40	250.2	8.5	177635	38	AC010654	AC010654 Homo sapi
41	246.4	8.3	2624	8	AB046644	AB046644 Macaca fa
42	245.4	8.3	1800	4	GGCKR1	X15558 Chicken ckr
43	245	8.3	179215	58	AC073422	AC073422 Homo sapi
44	242.2	8.2	138408	65	AC079555	AC079555 Mus muscu
45	238.2	8.1	175959	49	AC021154	AC021154 Homo sapi

## ALIGNMENTS

```

HTGERPAPCPDCKDSFSSOKSLNLTIRKSHIRDGAFCICAIGQITDDEBRLLAHOKKHD
V"
BASE COUNT      568 a      991 c      848 g      547 t
ORIGIN

Query Match      100.0%; Score 2954.. DB 37; Length 2954;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 cgcgtttgtccagagttcttcagagttgtcgttgcacgtccgagtgtgcatagaagaacagca 60
        |||
Db      1 CGCCTTTGTCGAGTCTTCAGATTCACAGATTCGCTGTGCACACTCGAGTGTGCATAGAAACACA 60

QY      61 gacacagggagaggggcagacataaaggcactgttagggagcagttgtgccacatttctycagag 120
        |||
Db      61 GACACAGGGAGAGGGGCACACATFAMGGCACTGTATGGAGACAGTGGCCACATTTCCTCAAGAG 120

QY      121 gaagaacagatctctgaagacgttcgtttgcaagggggcccccttggccatctgtggccagagcc 180
        |||
Db      121 GAAGAACCAGATCTGGAACGTCTGTTCAGAGGGGCCCCCTGTGGCCATGGCCTGGCCAGCCC 180

QY      181 cgactccctttctggccctcccaagagatccacccacagaccttggggaaagagatcccgcgagg 240
        |||

```

RESULT	1
LOCUS	AF201303
DEFINITION	Homo sapiens dhfr oribeta-binding protein RIP60 mRNA, complete cds.
ACCESSION	AF201303
VERSION	AF201303.1
KEYWORDS	GI:6716713
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2954) Houchens, C.R., Montigny, W., Zeltser, L., Dailey, L., Gilbert, J.M. and Heintz, N.H.
TITLE	The dhfr oribeta-binding protein RIP60 contains 15 zinc fingers: DNA binding and looping by the central three fingers and an associated proline-rich region
JOURNAL	Nucleic Acids Res. 28 (2), 570-581 (2000)
PMID	10606657
ERENCE	2 (bases 1 to 2954) Houchens, C.R., Gilbert, J.M. and Heintz, N.H.
UTHORS	Direct Submission
TITLE	Submitted (02-NOV-1999) Pathology, University of Vermont, Soule
JOURNAL	Medical Alumni Building, Burlington, VT 05405, USA
FEATURES	Location/Qualifiers
Source	1..2954
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_line="Hela"
	130..1833
	/note="contains 15 C2H2-Kruppel-like zinc fingers"
	/codon_start=1
	/product="dhfr oribeta-binding protein RIP60"
	/protein_id="AAF26712.1"
	/db_xref="GI:6716714"
	/translation="MLERRCGKLPLAMGLAOPRLTSGPSOESPTLAKESRGLROOGTSVAOSGAOPGRNACACRHRHFGMWALMLITRRCOARLPPECGSRFHAFELMLHROWHAATPDLGFACHLGSGSFGWALVLIHLAHSKAAQAPICGCEKREMRMGLRAHRLRRPAPAREPRTICGSGSPFMQOLVLAHVALEAPAAKRCRPPGSRPVTARPGSDANDRPFOCACCCRRRHNPNLIAHRRVHTGERPHOCPECGKFTNNPYLTSRHITGEKIPYCKEGRERFRHKNPLLSKTIHKRSYGQAAPGGSPLDPAEPTDSAEIRTPAVPLKPAOEPGPAPRPEDQDLEAPSLTSCDGSRSFLERFLRHARDHTEPERPTCAECGKNGKTKTHLVASHVSGERPEFACCEGGRFSQSHLAHRRPDHAPDPFPCPDGCAKFRHRYLAHRRRIITGEKPYVCPGCGAKAFQKSNLVSHRR
CDS	

[illegible]

QY	961	tatctgattctcgcaacgagcgatccacacacgagcgaaagcccttaccgtgtgaagagttgc	1020
Db	961	TATCTGATTCTCGCAACGAGCGGATCCACACCGGAGAAAGCCCTTACCCTGTGAAAGAGTC	1020
QY	1021	gaccgcgcctctccgagcaacaacccaacccctgtctcacaagcaaatctcacaagcgatcc	1080
Db	1021	gaccgcgcctctccgagcaacaacccaacccctgtctcacaagcaaatctcacaagcgatcc	1080
QY	1081	gagaggttcgagcccaagtcgcgccccgagcccgaggagaccccaagtctgcagccgcgccag	1140
Db	1081	GAGAGGTTCGAGCCCAAGTCGCGCCCGAGGCCCGAGGAGCCCCACCACTTCCAGCCGCGCCAG	1140
QY	1141	gagtcgcgcgagccgagcccaaccccgagcggttaactcttgaagaccgagccagagagcgccgca	1200
Db	1141	GAGTCGCGGAGCCGAGCCCAACCCCGAGCGGTTAACCTCTGAAACCGGCCCCAGAGACCCGCCCA	1200
QY	1201	ggagcccgccagagacaaccgcgacagaccgcatcgaaagcccccctccctctacagctgc	1260
Db	1201	GGGAGCCCGCCAGAGACACCCGCGAGGACCCGATTCGAAAGCCCCCTCCCTCTACAGCTGC	1260
QY	1261	gagacatcgcgagcgagaggtcttcgagtcgagcggtcttcctgcgagcccaagcggaagac	1320
Db	1261	GAGACATCGCGGAGGAGAGGTCTTCGAGTCGAGCGGTCTCTCTCGGCCCCACACGCGCACGAC	1320
QY	1321	accgggagagcgagccctctcaactctgcgcgagttgcggagaaagaaattctgcgcaagaagcgac	1380
Db	1321	ACCGGGAGAGCGGCGCTTCAACCTCTGCGCGAGTGGGGAGAAATTTGGGCAAGAACGCGAC	1380
QY	1381	ctgtgtgagcaatctgcgcggtgtgcaactcgcgagcgagcgagcccttcgcctgcgcgagagttgcgc	1440
Db	1381	CTGTGTGAGCAATCTGCGCGGTGTGCAACTCGCGAGCGAGCGGCGCTTCTGCGTGGAGAGTGGCGC	1440
QY	1441	cgcgctctccccaagagcaacatactgcggcgagcaatctgcgcgagcaagacgcccgcgcatcg	1500
Db	1441	CGCGCTCTCCCAAGAGCAACATACTGCGGCGAGCAATCTGCGCGAGCAAGACGCCCGCGATCGG	1500
QY	1501	cccttcgtgtgtccgcgactgcggaagagccttcgcgcacaaacccctacgtgcgcgagac	1560
Db	1501	CCCTCTGTGTGTCCGCGACTGCGGAAGAGCCTTCTGCGCAAAACCTTACTGTGCGCGCGAC	1560
QY	1561	cgaggagatacaagaccgaggaagagccctacgctcccccgcagatctgagaaagcccttaag	1620
Db	1561	CGAGGAGATACAAAGACCAGGAAGAGCCCTACGCTCCCCCGCAGATCTGAGAAAGCCTTACAGC	1620
QY	1621	cagaagtcacaacacttggtgtgtgcacgcgagcgatccacaagcgagcgagcccttaacgctgt	1680
Db	1621	CAGAAGTCACAACACTGGTGTGCAACGCGGCATTCACAAGGGGAGACGGCCCTTACGCGCTT	1680
QY	1681	ccgcgactgcgacacgcgagctcttcagccaagaaagtccaactcatacccaacgcgaagagccac	1740
Db	1681	CCCGACTGCGACACGCGAGCCTTCCAGCCAGAAAGTCCAACCTTATCAACCAACGCGAAGAGCCAC	1740
QY	1741	atccggagaaagcgccctctctgtctgtgcaactgttgcaagacccttgacagcaagagagaa	1800
Db	1741	ATCCGGGAAGCGCCCTCTCTGCTGTGCAACTGTGCAAGACCTTGACAGCAAGAGAGAGA	1800
QY	1801	ctctctgcccacaagaagaaagcaagatgtctgagaacggttgagcgaggcggtgtgtgcgtba	1860
Db	1801	CTCTCTGCCCAACAGAAAGCAAGATGTCTGAGAACGGGTGAGCGGCGCGCTGTGTGGCTGA	1860
QY	1861	gagagagctcgaggctcctctctgtgtgtgagatctgacatctgagcttbggggtgccttcgaatgc	1920
Db	1861	GAGAGAGCTCGAGGCTCCTCTCTGTGTGGAGATCTCGAAGTGGGCTGTGGGGTGCCTCTGAATGC	1920
QY	1921	tggagatagggagaaatggaatctctagaaagggagttgaaagatgctgagagcttgagcttggtg	1980
Db	1921	TGGAGATAGGGAGAAATGGAAATCTCTAGAAAGGGAGTTGAAAGATGCTGAGAGCTGAGCTGGTGTG	1980
QY	1981	ggccctctctagagagaggttcaaccgcggtgtgccaagagaaacacattccaagcgcaag	2040
Db	1981	GGCCCTCTCTAGAGAGAGGTTCAAACCCGCTGTGACCAAGGAACCCACTTCCAAAGCGCAAGG	2040

QY	2041	acgagggcctccagcgtggtgtgtgtcaagaagcccgctccgactgagcccttgcctctggaaa	2100
Db	2041	acgccgacctccagctggctgtgtgtgtcaaggctccgctccgactgagcccttgcctctggaaa	2100
QY	2101	agcagaatacacatccgccccctacagccccctcgtgctcagaaggagccacagctggaaagga	2160
Db	2101	agcagaataacatccgccccctacagccccctcgtgctcagaaggagccacagctggaaagga	2160
QY	2161	ggcctccatcctctggttalttaagcccttaagccccgtctcttactgttaagttaacttag	2220
Db	2161	ggcctccatcctcctggttalttaagcccttaagccccgtctcttactgttaagttaacttag	2220
QY	2221	atcatttttggagaagcagcgggtgtgttagagcctccgtataatgaatgatctgtggtcagatacag	2280
Db	2221	atcatttttggagaagcagcgggtgtgttagagcctccgtataatgaatgatctgtggtcagatacag	2280
QY	2281	cttgaggaacccgtgctgtgctctgttgaacagaacttggcctcttgcagcagaagaaggtgt	2340
Db	2281	cttgaggaacccgtgctgtgctctgttgaacagaacttggcctcttgcagcagaagaaggtgt	2340
QY	2341	aagcgaagccactctaacctctcccttccctcccaactcgtccccctcgtlaagcacaccag	2400
Db	2341	aagcgaagccactctaacctctcccttccctcccaactcgtccccctcgtlaagcacaccag	2400
QY	2401	acttggagaagaccctgtctgtctgttataacttccatctctctctctccccaagaagcagatc	2460
Db	2401	acttggagaagaccctgtctgtctgttataacttccatctctctctctccccaagaagcagatc	2460
QY	2461	ccaagagcatctaactccttgtctgtctcgtcttattctgtgcgccccctccagcagctgaag	2520
Db	2461	ccaagagcatttaactccttgtctgtctcgtcttattctgtgtgcgccccctccagcagctgaag	2520
QY	2521	cctcccccgtgctgtcagcagcaactgtgtccaggtctgtctgaacaccgcagccccctcc	2580
Db	2521	cctcccccgtgctgtcagcagcaactgtgtccaggtctgtctgaacaccgcagccccctcc	2580
QY	2581	ttcgtccctctccacagctcagaatgtcaagcagaagcagcgcgcgcatgtgtatgtgaagc	2640
Db	2581	ttcgtccctctccacagctcagaatgtcaagcagaagcagcgcgcgcatgtgtatgtgaagc	2640
QY	2641	cagctgaagggaagtgctgtcgtgtgagtttccctctctccatctctcagacatatgacaaccctgc	2700
Db	2641	cagctgaagggaagtgctgtcgtgtgagtttccctctctccatctctcagacatatgacaaccctgc	2700
QY	2701	ctctgctgtgaagcaactaagtgacaaggaacttccgcacctctcgaagccctgtgaatgtct	2760
Db	2701	ctctgctgtgaagcaactaagtgacaaggaacttccgcacctctcgaagccctgtgaatgtct	2760
QY	2761	aattgttgaagaatttcaatttgtttaagaatcccttccctataaagaatgaattctgcctttcc	2820
Db	2761	aattgttgaagaatttcaatttgtttaagaatcccttccctataaagaatgaattctgcctttcc	2820
QY	2821	tataattctacatattggccctgtctgtctctctcgtgaactaaacagacaaccaattta	2880
Db	2821	tataattctacatattggccctgtctgtctctctcgtgaactaaacagacaaccaaccaattta	2880
QY	2881	ccccctcttccaactagagaataaacaattgtgttttgaaaaaaataaaaaaataaaaaa	2940
Db	2881	ccccctcttccaactagagaataaacaattgtgttttgaaaaaaataaaaaaataaaaaa	2940
QY	2941	aaaaaaaaaaaaa 2954	
Db	2941	aaaaaaaaaaaaa 2954	

RESULT			PRI	
AC005586/c		DNA	30-SEP-2000	
Locus	AC005586	132150 bp		
DEFINITION	Hom sapiens PAC clone RP4-58AD14 from Yg31-q35,			
ACCSSION	AC005586			complete sequence
VERSION	AC005586.2	GI:8468933		
KEYWORDS	HTG.			
SOURCE	human.			

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 132150)

## AUTHORS

Sulston, J.E. and Waterston, R.

## TITLE

Toward a complete human genome sequence

## MEDLINE

Genome Res. 8(11):1097-1108 (1998)

## REFERENCE

2 (bases 1 to 132150)

## AUTHORS

Stonking, T., Ozerisky, P., Mohlmann, P. and Le, T.

## TITLE

The sequence of Homo sapiens PAC clone RP4-584D14

## JOURNAL

Unpublished

## REFERENCE

3 (bases 1 to 132150)

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (01-SEP-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## REFERENCE

4 (bases 1 to 132150)

## AUTHORS

Waterston, R.H.

## TITLE

Direct Submission

## JOURNAL

Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## REFERENCE

5 (bases 1 to 132150)

## AUTHORS

Waterston, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

## COMMENT

On Jun 12, 2000 this sequence version replaced gi:3907529.

## JOURNAL

Genome Center

## AUTHORS

Center: Washington University Genome Sequencing Center

## TITLE

Center code: WUGSC

## JOURNAL

Web site: http://genome.wustl.edu/gsc

## AUTHORS

Contact: sapiens@wustl.wustl.edu

## TITLE

Summary Statistics

## JOURNAL

Center project name: H\_DJ0584D14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mail to: [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

This clone was derived from human PAC library RPI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

## VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

## FEATURES

## Source

The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584D14; actual end is at base position 131954 of RP4-584D14.

The sequence RP4-584D14 from base position 21655 to 22775 is CT rich. This region was sized with PCR from genomic DNA and the hindIII digest with band size 3643 real, and 3642 insilico. The sequence RP4-584D14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.

## Location/Qualifiers

1..132150  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q31-q35"  
/clone="RP4-584D14"  
/clone\_id="RPI-4"  
230..315  
/rpt\_family="MIR"  
960..1288  
/rpt\_family="MER4-group"  
1360..1468  
/rpt\_family="L2"  
1562..1616  
/rpt\_family="L2"  
1733..2080  
/note="similar to EST AW510685 (NID:g7148763) hc89e08.x1"  
2051..2114  
/rpt\_family="L2"  
2148..3224  
/note="CpG island (MGC=70.7, o/e=0.90, #CpGs=107)"  
2347..2426  
/rpt\_family="CG-rich"  
3141..3232  
/rpt\_family="Alu"  
3235..3368  
/rpt\_family="Alu"  
3795..3888  
/rpt\_family="MIR"  
4037..4199  
/rpt\_family="Alu"  
4308..4593  
/rpt\_family="MER4-group"  
4872..5309  
/note="similar to EST A1827463 (NID:g5448134) wf29f03.x1"  
4882..4909  
/rpt\_family="AT-rich"  
5679..5773  
/rpt\_family="MIR"  
5947..5952  
/note="match to EST AA482420 (NID:g2210098) zt34f02.r1"  
6148..6450  
/rpt\_family="Alu"  
6453..6690  
/rpt\_family="MER4-group"  
6695..7007  
/rpt\_family="Alu"  
7008..7117  
/rpt\_family="MER4-group"  
7180..7335  
/rpt\_family="MER4-group"  
7615  
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"  
7883..8397  
/note="match to EST W72943 (NID:g1383235) zd54f12.r1"  
7886..8357  
/note="match to EST A1144211 (NID:g3666020) qb93g11.x1"  
7886..8562  
/note="similar to EST AW411281 (NID:g6936822) fh11d05.y1"  
7886..8403  
/note="similar to EST AW246540 (NID:g6589533)"  
7886..8451  
/note="match to EST A1129215 (NID:g1689084) zn84b02.s1"



```

STS
misc_feature 7886..8133 /db_xref="GI:1131717"
misc_feature 7886..8146 /note="similar to EST A1868891 (NID:g5542895) wc49f12.x1"
misc_feature 7888..8340 /note="match to EST H37790 (NID:g907289) yp46h05.s1"
misc_feature 7896..8372 /note="match to EST R72488 (NID:g846520) yj90g06.r1"
misc_feature 7976..8575 /note="match to EST R70860 (NID:g844377) y149h02.r1"
misc_feature 8008..8309 /note="match to EST AA335898 (NID:g1988160)"
misc_feature 8075..8451 /note="match to EST AA121027 (NID:g1678432) zn84g11.r1"
misc_feature 8192..8564 /note="match to EST AA404837 (NID:g6923894)"
misc_feature 8228..8362 /note="match to EST AA040559 (NID:g1516964) zf06d08.r1"
misc_feature 8289..8732 /note="match to EST AA843440 (NID:g2929958) ak07h10.s1"
misc_feature 8302..8928 /note="similar to EST AA059375 (NID:g1553199) zf66g07.r1"
misc_feature 8433..8746 /note="match to EST N45962 (NID:g187128) yj34h02.r1"
misc_feature 8440..8968 /note="match to EST AA482420 (NID:g2210098) zt34f02.r1"
misc_feature 8459..8461 /note="match to EST AA481966 (NID:g2209644) zv43e06.r1"
misc_feature 8463..8928 /note="match to EST AA053336 (NID:g1543923) zf60g12.r1"
misc_feature 8591..9167 /note="match to EST AA129342 (NID:g1689125) zn84p02.r1"
misc_feature 8616..9159 /note="match to EST AA481966 (NID:g2209644) zv43e06.r1"
misc_feature 8707..8845 /note="match to EST AA401523 (NID:g6920209)"
misc_feature 8743..8744 /note="match to EST R18324 (NID:g771934) yf99d11.r1"
misc_feature 8754..9069 /note="match to EST R18324 (NID:g771934) yf99d11.r1"
misc_feature 8776..9343 /note="match to EST N36800 (NID:g1157942) yj34h02.s1"
misc_feature 8826..9377 /note="match to EST A1359215 (NID:g4110836) qz26g11.x1"
misc_feature 8835..9229 /note="similar to Oryctolagus cuniculus EST C83534 (NID:g3062491)"
misc_feature 8841..9377
Query Match 92.4%; Score 2729; DB 9; Length 132150;
Best Local Similarity 99.1%; Pred. No. 0; Mismatches 20; Indels 5; Gaps 5;
Matches 2796; Conservative 0;

```

```

QY 407 gcttcgcacatgccccctcttcttgaactgacccgcccagctcgtctgcccccaag 466
DB 10406 GCTTGGCCATGCCCCCTTCTTAGCACTGCACCCGACAGTCCAGTCTGCCACCCAG 10347
QY 467 acctggccttgctcgcacacctctgtggagagagcttcagagcttgggtgcccgttc 526
DB 10346 ACCTGGCCTTGGCTGCGCACCTGCTGGGCAAGACTTCCAGAGCTGGGCTGCTTC 10287
QY 527 tgcattcgtggtcccatctgaactgcaagaaccccaatcgctttcccaatgagagac 586
DB 10286 TGCATCTGCGGGCCCATTCAGCTGCAAGGCGGCCCATCGCTTCCCAATGCAAGAGAC 10227
QY 587 gcttcgtgacgaagaacacttcagatcatctctggcggtgtgcaccccccgcgcg 646
DB 10226 GCTTGTGCGCACGAAGACACTTTCGACTATCTGCGGCGGTGCACTTCGCCGCCG 10167
QY 647 aggcgcgccttcaatctgacgaactgtgacgagcttgcacagtggaacagctag 706
DB 10166 AGGCCGCGCCTTCATATGCGGCAACTGTGCGGAGACTTGGCCAGTGGGACAGCTAG 10107
QY 707 ttgcccacaagcggtgtcaagtagcttagagcccttgagagagcgccgacagagcttg 766
DB 10106 TTGCCACAAGCGGgtgacagtagagcccttgagagagcgccgacagagcttg 10047
QY 767 ggcgcgcgcgaagggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 826
DB 10046 GGCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9987
QY 827 accgccccctccagctgtgctgtgtgtgcaagcgcttcgagcaaaccccaacttgatc 886
DB 9986 ACCGCCCCCTCCAGTGTGCTGTGTGCAAGCGCTTCCGACAAAGCCCAACTGTATG 9927
QY 887 ctcaacgcgcgtgtgcaagcggtgagagcgcccccagctgcccgaatgtgggaagcgct 946
DB 9926 CTCACCGCGCGTGCACACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9867
QY 947 ttacaaataagccctatctacttcgaccgagcgtacacaaacggagagaagccctacc 1006
DB 9866 TTACCAATAAGCCCTATCTGACTTTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 9807
QY 1007 cgtgcaaaagctgagcgcgccgcttcgagcaaaccccaactgtgtctcaagcaaga 1066
DB 9806 CGTGCAAAGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9747
QY 1067 ttcaagaagatccagagaggtctggccagagcgcccccgcgcgcgcgcgcgcgcgcgc 1126
DB 9746 TTCAAGAAGATCCGAGAGGgtgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 9687
QY 1127 caagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1186
DB 9686 CAGCCGCGCCCCCAGAGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9627
QY 1187 agagagcgcgccagagggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1246
DB 9626 AGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9567
QY 1247 cccctacagctcgcagcaactgtgagagagcttcggtcgtgagagcttcctcgtggcc 1306
DB 9566 CCTCTTACACTGTGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9507
QY 1307 accagcgagcagcaacacagggagagcgcccttcaactctgcgcgaatgtggggaagaactcg 1366
DB 9506 ACCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9447
QY 1367 gcaagaagaagcactgtgtgagcacttcgagctgcaacttcgagcgagcgcccttcgct 1426
DB 9446 GCAAGAAGAAGCACTGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9387
QY 1427 gcaagaagatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1486
DB 9386 GCGAGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9327
QY 1487 acgcgccgcgctgcgccttcgtgtgtccgcagctgagcgagccttcgcacaaaccc 1546

```

D	b	9326	ACGGCCCGGATCGGCGCTCTGTGTGTGCCAGTGGCGCAAGGCGCTTCCGCCAACAACCT	9267
O	y	1547	acctggcgcggaacacggcgcatccaacccggcgagaagccctacgtctgtgccgactgcg	1606
D	b	9266	ACCTGGGGGGCCACCGGGCGGATCCACACCGGGAGAAAGCCCTACGTCCTCCCGACTCG	9207
O	y	1607	gcaagcccttaagccagaatgccaaactgtgtgcacccggcgcatctcacaagggcgagc	1666
D	b	9206	GCAAGCGCTTCAGCCACAAGTCCACACTGGTGTGCAACCGCGGCATCCACAGGGCGAGC	9144
O	y	1667	ggccctacgcctgtcccactgcgaccgcagcttcaagccagaagtccaactcaacc	1726
D	b	9146	GGCCCTACGGCTGTGCCAGCTGGGACCGCGAGCTTCAGCCAGAAGTCCAACTCATCACCC	9087
O	y	1727	accgcaagagccacatcccggaagcgcccttcgcgtgcacatgtgtggcgaaccttcg	1786
D	b	9086	ACCCGAAGAGCCACATCCGGAGCGGCCCTTGTGCTGTGCACTGTGTGGCCAGACCTTCG	9027
O	y	1787	acgaagagagagaagactcttcgtccacccagaagaagacagatctcgaacggtgtggcgcg	1848
D	b	9026	ACGACGAGAGAGACTCTCTGGCCACACAGAAGAAGCAGCTGTGTGAGACGGTGGGGCGG	8967
O	y	1847	gcgctgtgtgcctgagagaagggctcgtgtcctgtgtgtgtgaatgcagttgcgtggg	1906
D	b	8966	GCCCTGTGTGCTGTGAAGAGGGCTGTGGGCTCTTGTGTGTGGAGTCCACTGTGGCTGGGGG	8907
O	y	1907	tgccctgcctagtgtgtgagtagaggagacaatgggaatccctagaagggaatgaatgtgcgg	1966
D	b	8906	TGCTCGCTAOTGCTGGAAGTAGGGGACAAATGGGAATCTTAGAGGGATGTGAAGACGGGG	8847
O	y	1967	gaatgaagcttgggttggccctgtctagcgaagagggtcaaccgccgtgtgcaggagaaaccaac	2026
D	b	8846	GAGTAGACTGTGGTGGGCCCTGCTAGCGAGAGAGGTCAACCCGGTGGCCAGGAGAACCCAC	8787
O	y	2027	ttccaagcggaaggagacgcggccttcagcgtgtgtgtgtaagcttcogtctgtactgc	2086
D	b	8786	TTCCAAAGGCGAGGAGCCCGGCTCCACGCTGGTGTGTGTAAGGCTCGTCTGTGACTGC	8727
O	y	2087	ctgtgcccttggaaaaagagaata-catccgcccccttaagccctctgtcctagaagagacc	2145
D	b	8726	CTGTGCCCTTGAAAAAGCAGCAATAGCATCGGCCCTTAGAGCCCTGTGCTGAGAGACC	8667
O	y	2146	accagttgaaaagaaag--ccctcatcctctgtgtatataagccttaatgcctctgctctt	2204
D	b	8666	ACCAGTGAAAAGAAAGACCTCCATCCTCTGTGATTAAAGCCTTAATGCCCTGTGCTTTT	8607
O	y	2205	acttaagttactc--aatacatctttggaagcgcggtgtgtagagtcctgttaatgaaatg	2263
D	b	8606	ACTGTAAAGTTACTTAAGATCATTTTGGGAAGCAGGGGTGTGAGTGCCTGTAAATGAATG	8547
O	y	2264	ctctgtgctatagatacagcttggagaacctgtgcctgtttagaacaaacttggcctt	2323
D	b	8546	CTCTGGGCTGTGATTCACCTTGTGGAAACCTGTGCTGGCTGTGTAGACAGCACTTGGGCTTTT	8487
O	y	2324	gcacgacgaagaaggtbaagcgaaagcaactactcctccctccctcccaactggcc	2387
D	b	8486	GCACGACGACAAAGGTTAAAGCGAACCTCTTAACCTGTCCCTTCCCTCCACACTGGCC	8427
O	y	2384	ccgtgcgtagaagcaaccagaacttggagaacccgctgtcgttataacttcaatcccttcc	2443
D	b	8426	CCTGCTGTAAGGACCCAGACATTGGAGAGACCCGCTGCTGTAAATACTTCCATCCTTCC	8367
O	y	2444	ttccccaagagcagatctccaagagcattactcctgtgtcctgtcctgttactctgtgcg	2503
D	b	8366	TTTCCCAAAGAGCAATCCCAAAGGCAATTTCCTCTTGTGTGCTGTGCTTAATCTGTGCC	8307
O	y	2504	cctccacgacgtgagagcccccctgtgctgtacagagaaactgttccaagcctctgtctg	2563
D	b	8306	CCTCCACGCGGTAGACCTCCCTCTGGCTGTACAGCAGCACTGTGTTCAGGCTCTGTGTG	8247
O	y	2564	aaacacgcagccctcctctgtcctcttcaacagctcaagatgtaagcaagaaagactgcg	2623

Db	8246	AACGCGAGACCCTCCTTGGTCGCTTCCAGACTCAGATGTCAACGGGAAGACTGCCG	8187
Oy	2624	cattgttgatggagggccacgctcgaagggaagtgtcgctgatcttccc-tttctccatttc	2682
Db	8186	CATTGGTGATGGAGGGCCACCTAAGGGGAAAGTTGGTGGAATTTTCCTTTCATTTC	8127
Oy	2683	tagcatat-gaacacctgcctcctgcctcttgagaacttagtgatgaaaggaattccgaccctc	2741
Db	8126	TAGCATATTGAGCACTGGCCCTCTGCCTTAGACATTATGGTGAAGAAGACTTCGCACTCCC	8067
Oy	2742	ttaggccctggatgatctcaatttgtttaagaattctaatttgtttaagaadtcctctataa	2801
Db	8066	TGAGCCCTCGATGATCTTAATTGTTAGAATTCATAATTGTAGAAAACCTCTCTTATAA	8007
Oy	2802	tgaatgaattctctcttccaataattctaacattggggccttgctctgcctcggagac	2861
Db	8006	TGAATGAATTCCTGCTTCCATATATTCTACCTATTGGGCTGTGTCGTCTCTGGAAC	7947
Oy	2862	taaacagaacaacattaccctccctlttccaactagaagaataacatttgtttlaga	2921
Db	7946	TAAACAACAACACATTACCCTCCTTTTCAAAC TAGAATAAAGATTGGTTTTAGA	7887
Oy	2922	a 2922	
Db	7886	A 7886	
RESULT	3		
LOCUS	AK025356		
DEFINITION	AK025356	2878 bp mRNA	PRT 29-SEP-2000
ACCESSION	AK025356	Homo sapiens cDNA: FLJ21703 fis, clone COL09895, highly similar to HSA245553 Homo sapiens mRNA for AP4 zinc finger protein.	
VERSION	AK025356		
KEYWORDS	AK025356.1 GI:10437856	oligo capping; fis (full insert sequence).	
SOURCE	Homo sapiens colon cDNA to mRNA, clone_lib.COL clone:COL09895.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (sites)		
TITLE	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
JOURNAL	NEDO human cDNA sequencing project unpublished (2000)		
REFERENCE	2 (bases 1 to 2878)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirakane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan		
COMMENT	(E-mail:cdna@ems.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
FEATURES	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
SOURCE	Location/Qualifiers		
	1..2878		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="COL09895"		
	/clone_lib="COL"		
	/rname="colon"		
	/rsize_type="colon"		
	/note="cloning vector pME18SF13"		
	1..2878		
	/note="highly similar to HSA245553 Homo sapiens mRNA for AP4 zinc finger protein"		

BASE COUNT 560 a 974 c 826 g 518 t  
ORIGIN

Query Match 92.1%; Score 2719.4; DB 37; Length 2878;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 2806; Conservative 0; Mismatches 31; Indels 8; Gaps 6;

QY 115 gacgagaaagaaacgagatgagagacgtctgtgagagggcccccgtgacgagcgtgac 174  
DB 13 GCGAGAGAAAGACGAGAGTGGAAAGTGTGAGAGGGGGCCCCGCGCATGGGGCCCCG 72

QY 175 cagcccgactcccttcctgagccctcccaagagtlcaaccagaccctggggaaagatcc 234  
DB 73 CAGCCCGAGTCTCTTGTGGGGCCCTCCAGAGTCAACCCAGACCCTGGGGAAAGATCC 132

QY 235 cgcgggctgagcaacaagacgtcagtgagccagtcgtgtgcccagcccaagcagcag 294  
DB 133 CCGGGGCTGAGGCAACAGGACAGTCAAGTGGCCAGTCTGGTGGCCCAAGCCCCAGGAG 192

QY 295 gcccatcgctgagccactgagagacatccctgagtggtgtgtgtgtgtgtgtgtgtgt 354  
DB 193 GCCCATGCTGTGCTCCACTGTCAAGGCACTTCCTGGTGGGGTGGCTGTGGCTTTCAC 252

QY 355 accgcgcgtgagcagggccgcgtgacctgacctgacctgagtggtgagcgtgacctgac 414  
DB 253 ACCGCGCGGTGCGAGGCGCGGCTGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTG 312

QY 415 catgccccttcttaagactgacagcgagacgtcagtcgtgtgtgtgtgtgtgtgtgtgt 474  
DB 313 CATGCTCCCTTCTTACGACTGACACGCGCAGGTCACTGCTGCTGCTGCTGCTGCTG 372

QY 475 ttgtgctgacactctgt 534  
DB 373 TTTGCTGCTGCTGCTGTGTGGGAGAGCTTCCGAGGCTGGTGGCTGCTGCTGCTGCTG 432

QY 535 ctggcccatcagctgaaagcaaccatcgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 594  
DB 433 CCGGCCCATTCACGTGCAAAAGCGGCCCATGCTGTGTGTGTGTGTGTGTGTGTGTGT 492

QY 595 cgaagaaagacgtctgagactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 654  
DB 493 CGAGAAAGACGCTGAGCTCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 552

QY 655 cccttcataatgagcaactgt 714  
DB 553 CCTTCAATATGCGGCAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 612

QY 715 aagcgggtgagcgttaactgagggccctgagagagcgacgaaggtctgtgtgtgtgtgt 774  
DB 613 AACGGGTGCACTAGCGGAGGCGGCTGTGAGAGAGGCGGAG---AGGCTGTGTGTGTGT 669

QY 775 cccaggggacggcccgctgtgacggcccccggcccgctgtgtgtgtgtgtgtgtgtgtgt 834  
DB 670 CCCAGGGGCGCCCGCGGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 729

QY 835 ttcaatgtgtcgt 894  
DB 730 TTTCAAGTGTCTGT 789

QY 895 cgcgtgacacgggagcaggggcccacacagtgcccaggtgtgtgtgtgtgtgtgtgtgt 954  
DB 790 CCGGTGACACGGGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 849

QY 955 aagccatctgacttgcacggcgagatccacacggcgagagaccccttaccgtgtgacaa 1014  
DB 850 AAGCCATATCTGACTTGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 909

QY 1015 gagtgtgagcgcgcttcgagcaaaacccaactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1074  
DB 910 GAGTGTGGGCGCGGCTTCCGCGCAAAACCACTGTGTGTGTGTGTGTGTGTGTGTGTGT 969

QY 1075 cgttcagaggt 1134

DB 970 CGATCCGAGGGGTGCGGCCAGGCGGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1029

QY 1135 ccccaagatgc 1194  
DB 1030 CCCAGAGAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1089

QY 1195 ccgcgaagggcccgccagagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1254  
DB 1090 CCGCAGGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1149

QY 1255 agctgcagcagctgc 1314  
DB 1150 AGCTGCGACGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1209

QY 1315 caagcaacggggagcgcccttcaactgagcgagtgcgggagaaacttcgcaagaag 1374  
DB 1210 CAGCAGACCGGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1269

QY 1375 acgcacctgt 1434  
DB 1270 ACGCAGCTGT 1329

QY 1435 tgcggcgccgcttctccagggagcagcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1494  
DB 1330 TGGGCGCGCGGCTTCTCCAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1389

QY 1495 gaticgcctcgt 1554  
DB 1390 GATCGGCGCTTGT 1449

QY 1555 cggacacggcgatccacacggcgagaaagcccttgcgtgtgtgtgtgtgtgtgtgtgt 1614  
DB 1450 GCGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1509

QY 1615 ttcaagcaagatccaaactgt 1674  
DB 1510 TTTCAAGCAAGATCCAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1529

QY 1675 gctgtccagactgagcagcagcttcaagcagaagtcacacatcaaccacgcgcaag 1734  
DB 1570 GCGTGTCCCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1629

QY 1735 agccacatccggagagcgcccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1794  
DB 1630 AGCCACATCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1689

QY 1795 gaggactcttgc 1854  
DB 1690 GAGAGACTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1749

QY 1855 ggtcgaagagaggt 1914  
DB 1750 GGTGAGAGAGAGGCTGT 1809

QY 1915 taagt 1974  
DB 1810 TAAGT 1869

QY 1975 ttgt 2034  
DB 1870 TGGGT 1929

QY 2035 gcaaggagc 2094  
DB 1930 GCAAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1989

QY 2095 ttgaaagagcaata-catccgcaccttaagagccctctgtgtgtgtgtgtgtgtgtgtgt 2153  
DB 1990 TGGAAAGAGCAATTAACATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2049

QY 2154 aaaggaag-cccttcacatccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2212

Db 2050 AAGAGAGACCTCCATCCTGTGTATTAACGCCCTTAATGCCCTGTCTTTACTGTAG 2109  
Qy 2213 ttaact-agaatcatttttgaagcagcgctgttagatctctgtaaatgaatgctctgagc 2271  
Db 2110 TTACTTAGATCATTTTGGAGGAGGCGCTGTAGAGTCTCTTAATGATGCTCTGGGC 2169  
Qy 2272 tagatagagcttgggaacctgtctgtctgttagaacaagaacttggccttggcagcag 2331  
Db 2170 TAGATACAGCTTGGAGAACCTGTGCGCTTGTAGAACACACTTGGCGCTTGGCAGCAG 2229  
Qy 2332 caagaggtgaagcgaagcgaacttaactctctctctctctctctctctctctctctct 2391  
Db 2230 CAAGAGGTGAAGCGAAGCGACCTTACCTCTCCCTCCCTCCACCTGCGCCCTGCGTA 2289  
Qy 2392 ggcacccgaacttggagagacccgctgtctgttaataacttccactctctctctctctct 2451  
Db 2290 GGCACCCGACACTTGGAGAGACCCGCTGTGTAGTACTTCCATCTCTCTCTCCAAA 2349  
Qy 2452 gacgagatcccaaggaacttaactctctgtctgtctgtctgtctgtctgtctgtctgt 2511  
Db 2350 GAGCAGACCCCAAGGCAATTTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2409  
Qy 2512 cgtctagagcctccctctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtct 2571  
Db 2410 CGCTAGAGACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2469  
Qy 2572 agccct 2631  
Db 2470 AGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2529  
Qy 2632 atggaagggcagctgaaggggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2690  
Db 2530 ATGAGAGGCGACACTGAGGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2589  
Qy 2691 -gacacccgtgctctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2749  
Db 2590 GACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2649  
Qy 2750 tggatgatcttaattgttagaattcctaattgttagaattcctaattgttagaattcctaatt 2809  
Db 2650 TGGATGATTTCTAATTTGTAAGAATTTCTAATTTGTAAGAATTTCTAATTTGTAAGAAT 2709  
Qy 2810 ttctgtct 2869  
Db 2710 TTCT 2769  
Qy 2870 acaacatttaccct 2929  
Db 2770 ACAACCATTTTACCT 2829  
Qy 2930 aaaaaaaaaaaaaaaaaaaaaa 2954  
Db 2830 AAAAAAAAAAAAAAAAAAAAAA 2854

RESULT 4  
LOCUS HSA245553 2878 bp mRNA PRI 18-AUG-1999  
DEFINITION Homo sapiens mRNA for AP4 zinc finger protein.  
ACCESSION AJ245553  
VERSION AJ245553.1 GI:5748564  
KEYWORDS AP4 gene; AP4 protein; zinc finger protein.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2878)  
AUTHORS Dobner,T.G., Fischer,M. and Grothl,P.  
TITLE Cloning of a novel zinc finger protein  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2878)  
AUTHORS Dobner,T.G.  
TITLE Direct Submission

JOURNAL Submitted (17-AUG-1999) Dobner T.G., Institut fuer Medizinische  
Mikrobiologie und Hygiene, Universitaet Regensburg,  
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY  
FEATURES  
Source  
1. 2878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Hela"  
gene  
106. 1776  
/gene="AP4"  
106. 1776  
/note="zinc finger protein"  
/product="AP4 protein"  
/protein\_id="CA53100.1"  
/db\_xref="GI:5748565"  
CDS  
BASE COUNT 528 a 973 c 839 g 538 t  
ORIGIN  
Query Match 89.9%; Score 2654.6; DB 77; Length 2878;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 2772; Conservative 0; Mismatches 49; Indels 8; Gaps 8;  
Qy 107 catltctcgaagagaagaacgatgtctgtgaagcgtctgtcgaagggcccccgtgcatgg 166  
Db 51 CACCTCAGCAGAGAGAAACCGATGCTGGAAGCTGTCGCA-GGGCCCTGCGCATGG 109  
Qy 167 gcttgccagccagccagcct 226  
Db 110 GCCCGCCAGCCCGACTCTTCTGCGCTCTCCAGAGATCACCCAGACCTGGGGA 169  
Qy 227 aggaatcccgcggtgtgagcaagaagcagctcagtgccagctgtgtgccaaagcc 286  
Db 170 AGGAGTCCCGCGGCTGAGAGCAACAAGCAGTCAGTGCCCCAGTGTGCCAAGGCC 229  
Qy 287 caggcagggcccatcgtctgtcccaactgtcgaaggaacttccctgtgtgtgtgtgtgt 346  
Db 230 CAGGCAAGGCCCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 289  
Qy 347 ggtctcaaccccggtgtgagccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 406  
Db 290 GGGCTTACACCCCGCGGTGCTGAGGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 349  
Qy 407 gcttcgcacatgccct 466  
Db 350 GCTTTCGCATGCCCTCTTACGACAGTGCACGGCAGGTCCATCTGCTGCGACCCAG 409  
Qy 467 accctggcttgcctgtcgaacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 526  
Db 410 ACCTGGGCTTTCCTGTCACACTGTGTGGGAGAGCTTCCAGAGCTGGGTGCGCTGCTTC 469  
Qy 527 tgcattctgtgcccattcagctgtgaagaacacccatgtctgtgtccaaatgtgagaac 586  
Db 470 TGCATCTGGGGCCCATTCAGTGAAGGGCCCATGCTGCTTGTCCAAATGCGAGAGAC 529  
Qy 587 gcttcgcgaagaagaagcgtcagctcattcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 646  
Db 530 GCTTCTGGCGAGGAAGCAGCTTCAAGCTCACTGACAGCGGTGCCACCTCCGCGCCGG 589  
Qy 647 agcccgcccttcatatgtcgaacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 706

D	b	590	CGGCGCGCCCTTCATATGCGGAAACTGTGTGGCCGAGACTTTTGCCATGTGGACACAGCTAG	643
Q	y	707	ttgccccaagcgggtgtacgcttagcgtgagcccttgtagggacgcgaagcgaagcgtctg	766
D	b	650	TTGGCCCAAGCGGCTGCAGCTGTAGAGCCCTGGAGAGGCCGAGCAAGCTTCTGG	709
Q	y	767	ggccccggcccaaggcgccgcgcgggtgacgcgcgccccggcgccgggtggagatgcctcg	826
D	b	710	GGCCCCGGGTCCAGGGGGGCGCCCGCGGTGACCCGCCCCCGCGCGGTGAGATGTGCGTGC	769
Q	y	827	acgcgcccttcagatgtgacctgtgtgtggaaagcgtcttcggacaagcccaactgtatcg	886
D	b	770	ACCGCCCTTTCAGTGTGCTGTGTGTGTGGCAAGCGCTTCGCGCACAAGCCCATCTTGATGC	829
Q	y	887	ctcacgcgcgcgtgtacaaacggcgagcgcgcaccccaacagctgcgccagatgtgcggaaagcgt	946
D	b	830	CTCACGCCGCGGTGCACACGGGGGAGCGGGCCACCAAGTGCCTCGAGTGTGGGAGGCGCT	889
Q	y	947	ttaccaataagccctatctgaattgcacccgycgcatctcaacacggcggagaagcccttgc	1006
D	b	890	TTACCAATTAAGCCCTATCTACTTCTTCACCGGGGATTCACACCGCGCAGAAGCCCTTACC	949
Q	y	1007	cggtcaaaaggtgtgcgccgcgcgtcttcgcgcacaacaaacaaactgtgtgtctcaacaga	1066
D	b	950	CGTCCAAAGAGTGTGGCGCCGCCCTTCGCGCACAAACCAACTGCTGTCTACACACAAGA	1009
Q	y	1067	ttcaacaagcgatccgagggtgtcgcccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1126
D	b	1010	TTTCACAAGCGATCGAGGGGTGCGCCACAGCGCGCCCGCGCGGGGAGCCCCCAGCTGC	1069
Q	y	1127	cagccggcccccaaggagttccgcggcccgagcccaaccccggcggtacctgtgaaacgcgc	1186
D	b	1070	CAGCGCGCCCCCAGAGATCCGCGCGCGAGCCACCCCGCGGTACTCTGAAACCGGCGCC	1129
Q	y	1187	agagagccgc	1246
D	b	1130	AGGAGCGCGCCGACAGGGGCCCGCCGCAAGAACCCGCAAGAACCGATCGAAGCCCCCTCT	1189
Q	y	1247	ccctctacagctgagacgagactctcgagaggagcttccgcgtgtgagcgtctctctgcgcgc	1306
D	b	1190	CCCTCTACAGTGTGGAGACGACTGTGGGAGGAGCTTCCGCTGGAGGCTTCTCGCGGGCC	1249
Q	y	1307	accagcggcaacgacacccggggagcgccttcaactctgcgcgagtgcgggaaagaaactcg	1366
D	b	1250	ACCGCGCGGACAGCACCGCGGGAGCGGCCCTTCACTCGCGCCGACGAGTGGGAAAGACTTGC	1309
Q	y	1367	gcaagaagaagcaactgtgtgtgscgatctgcgtgtgactccggcgcgagcggcccttcgct	1426
D	b	1310	GCAAGAAAGACGACCTGTGTGTGGCCACTCGCGCGTGCACATCGGACAGCGGCCCTTTCGCT	1369
Q	y	1427	gcgagaggtgtcgcgcgcgcgtcttcccaaggcagccatcttgccgcgcgatctgcgcgcgc	1486
D	b	1370	GCAGAGAGTGTGGCGCGCGGCTTCTCCAGGGAGGCCATGTGGCGGGGCAATGGCGCGAC	1429
Q	y	1487	acgcgcccgatcggccctctgtgtgtccgaactgtgcgaagccttccgcacaacacct	1546
D	b	1430	ACGCGCCCGATCGGCTTGTGTGTGTCCGACTGCGGCAAGGCGCTTCCGCAACAACCT	1489
Q	y	1547	accgcggcgagcaacgcgcgcgatctcaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1606
D	b	1490	ACCTGGGGGCGCACCGCGGATCCACACCGCGGCAGAGGCGCTTACGCTGTGCCCGCATCTGG	1549
Q	y	1607	gcaagccttcagccaanaatccaactgtgtgtcgaacgcgcgcgcgcgcgcgcgcgcgcgcgc	1666
D	b	1550	GCAAGCGCTTCAAGCANAAATCCAACTGTGTGTGCGACCGGGGCAATCCACAGGGCGAGC	1609
Q	y	1667	ggccctaagcctgtccgc	1726
D	b	1610	GGTCTCTACGCGCTGTCCGACTGTGCAGCCGACGCTTACGACAGAACTCAACCTCATCACCC	1669
Q	y	1727	acgcgaagcgaacatctcggaagcgcgcgccttctgcctgtgtcatctgtgtcgcaagacctcg	1786
D	b	1670	ACCGCAAGAGCACAATCCGGAGGGCGCTTGTGCTGTGCATCTGTGTGGCAGACTTTCG	1729

OY	1787	acagcagagagagagcctctctgcccacacagaaagacagatgctcgaagcggctggcgag	1846
Db	1730	ACGACGAGGAGAGACTCTCTGGCCACCAGAGAGACGACATCTTGAAACGCTGAGCGG	1789
OY	1847	gcccgtgttggctctgagaaagggcctggggtccttcgttggtggagtcgaagtggcctggcg	1906
Db	1790	GCCGTGTTGGCTAGAGAGAGGCGCTGGGGTCTTCTGTGTGTGGAGTGGCAGTGGGCTGGGG	1849
OY	1907	tgctgacctagctgctgagtagtggggaacatctggaaacctagaagggaatgaaatgacgg	1966
Db	1850	TGCCCTGCCTATGTGCTGGAGTAGGAGGACAAATGGGAATCTTAGAGGGGATGGAAGACGGCG	1909
OY	1967	gaatgagactgggtggggccctgcctagcgaagaaagagtgtaaccccggtggccaggaaacccac	2026
Db	1910	GAGTGAGCTGTGGTGGCGCCCTGTAGGGAAGAGAGTCAACCCCGGTGGCAGGGAAGCCAC	1969
OY	2027	ttccaagcagcagga -ggccggcctcagaatggtgtgtgtctaaagctcgtccctgaatgc	2085
Db	1970	TTCCAAAGCGAGGAGCGCGGGCTCCAGCTGGTGTGTCTTAAGCTCCGTCCTGACTGC	2029
OY	2086	ccctgtgccctggaaaagcaagcaata -catccgccccttcacagcccctcgtgctagaaggagc	2144
Db	2030	CCTGTGCCCTGAAAGACAGCAATAGCATCCGCCCTTAGAGCCCTGTGGGTAGAGAGC	2089
OY	2145	caccagtggaaaggaag -ccctccatcctctggtatlaagccttaatgcccctgctctt	2203
Db	2090	CACCAAGTGGAAAGGAAACCCCTCCATCCTGTGATTAAAGCCCTTAAGCCCTGTCTT	2149
OY	2204	tactgtaagtacttagatactttttggaaagcagcggtgtgtagatctctgttaatlgatg	2263
Db	2150	TACTGTAGTACTTAATGATCATTTTTTGGAGCAGGCGTGGTAGACTCTGTAAATGATG	2209
OY	2264	ctctgggctagatatacaactctggagaacccctggccctgttagaagaacttggccctt	2323
Db	2210	CTCTGGGCTGATATACACTGTTGAGAAACCTGTGGCTTGTTAGACAGCACTGGGGCTTT	2269
OY	2324	ggcagcagcaagaggtgaagcgaagcgaactctacctctccctcccaacctggcc	2383
Db	2270	GCCAGCAGCAAGAGGTGAACGGAAGCGACACTTACTCTCCCTCCCTCCACCTGGCC	2329
OY	2384	cctgtctaggacaccacagacttggaaagaccgtctgctgttaatacttcaactctctctcc	2443
Db	2330	CCTGTGTAGGCAACCAACTTGGAGAGACCCGCTGTCTTAAATCTTCATCCCTTCC	2389
OY	2444	ttccaagaagcagatcccaagacatlaactcctt -ggtctgtccgcttatctctgcgc	2502
Db	2390	TTCCCAAGAGACAGATATCCCAAGCAATTTACTCTTGGGTCTGTCTCGCTTATCTGTGC	2449
OY	2503	ccctcccaagcgtctgagacacctccctcctgctgtctgaagcagactgtg -tccaagctctgtc	2561
Db	2450	CCCTCCAGCGCTAGAGACTCCCTCCCTGGCGTGTACAGACACTGTGTCCAGGCTCTTGTCTC	2509
OY	2552	tgaacaccgagccctcctctctgctctctccacaagctcaagcatgtcaagcgaaggaatgc	2621
Db	2510	TGAACACCGCAGCCCTCTCTTCCCTCTTCCTTCACAGCTCAGCATGTCAAGCAAGAGCTGC	2569
OY	2622	cgcaatgtatagtagggcagcagcttggagggaagtctggttgagtcttcc -tttcccaat	2680
Db	2570	CGCATTTGGTGTATGAGGGGCGACGCTGAGGGGAAGTtctgtagtTCCTTTCTCCATT	2629
OY	2681	tctagcatat -gaaacctgacctctgctctgtgagcaactttagtacaagaaactccgaact	2739
Db	2630	TCTAGCATATGAGACACTGGCGCTGTGCTTAGAGACTTAGCTCAGAGAACTTCGCAACT	2689
OY	2740	ccctgagccctgagatgtaattgttagaaattcctaattgtttagaaatcccttccat	2799
Db	2690	CCTAGAGCCCTGAGTGTCTTAATTTGTAGAAATCTTAATTTGTGAATCTCTTCCTAT	2749
OY	2800	aatgaatgaattcgtcttctctaataattctcaactatgggcctgtgtcttctctgga	2859
Db	2750	AATGAATGAATTCGTCTTCTTAATTTCTTAATTTCTTAATTTGGGCTGTGTCTCTCTGGA	2809

QY	2860	actaaacgaacacattaccctccctttcaacitgagagataaacatttgctta	2919
Db	2810	actaaacgaacacacattaccctccctttcaacitgagagataaacatttgcttt	2869
QY	2920	gaaaaaaa	2928
Db	2870	agaaaaaaaa	2878

RESULT	5			
AC073111				
LOCUS	AC073111	179225 bp	DNA	HTG
DEFINITION	Homo sapiens chromosome 7 clone RP11-511P7, WORKING DRAFT SEQUENCE			15-JUL-2000
ACCESSION	AC073111			
VERSION	AC073111.3	GI:9211463		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	Waterston, R.H.			
TITLE	The sequence of Homo sapiens clone			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 179225)			
AUTHORS	Waterston, R.H.			
TITLE	Direct Submission			
JOURNAL	Submitted (08-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA			
COMMENT	On Jul 15, 2000 this sequence version replaced gi:8572520.			

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH051IP07
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160317 bases at least Q40
Consensus quality: 166937 bases at least Q30
Consensus quality: 169751 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 176225; sum-of-contigs
Quality coverage: 4.24 in Q20 bases; agarose-fp
Quality coverage: 4.40 in Q20 bases; sum-of-contigs

*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1315: contig of 1315 bp in length
* 1316 1415: gap of unknown length
* 1416 2954: contig of 1539 bp in length
* 2955 3054: gap of unknown length
* 3055 4541: contig of 1487 bp in length
* 4542 4641: gap of unknown length
* 4642 6249: contig of 1608 bp in length
* 6250 8039: gap of unknown length
* 8039 contig of 1690 bp in length
* 8139: gap of unknown length
* 8140 9473: contig of 1334 bp in length
* 8140 9474 9573: gap of unknown length

```

*	9574	11903:	contig of 2330 bp in length
*	11904	12003:	gap of unknown length
*	12004	14193:	contig of 2190 bp in length
*	14194	14293:	gap of unknown length
*	14294	17408:	contig of 3115 bp in length
*	17409	17508:	gap of unknown length
*	17509	20169:	contig of 2661 bp in length
*	20170	20269:	gap of unknown length
*	20270	22745:	contig of 2476 bp in length
*	22746	22845:	gap of unknown length
*	22846	26671:	contig of 3826 bp in length
*	26672	26771:	gap of unknown length
*	26772	30438:	contig of 3667 bp in length
*	30439	30538:	gap of unknown length
*	30539	33209:	contig of 2671 bp in length
*	33210	33309:	gap of unknown length
*	33310	36685:	contig of 3576 bp in length
*	36686	36985:	gap of unknown length
*	36986	41449:	contig of 4464 bp in length
*	41450	41549:	gap of unknown length
*	41550	45860:	contig of 4311 bp in length
*	45861	45960:	gap of unknown length
*	45961	49359:	contig of 3399 bp in length
*	49360	49459:	gap of unknown length
*	49460	54617:	contig of 5158 bp in length
*	54618	54717:	gap of unknown length
*	54718	61279:	contig of 6562 bp in length
*	61280	61379:	gap of unknown length
*	61380	66606:	contig of 5227 bp in length
*	66607	66706:	gap of unknown length
*	66707	73439:	contig of 6733 bp in length
*	73440	73539:	gap of unknown length
*	73540	80804:	contig of 7265 bp in length
*	80805	80904:	gap of unknown length
*	80905	87301:	contig of 6397 bp in length
*	87302	87401:	gap of unknown length
*	87402	95200:	contig of 7799 bp in length
*	95201	95300:	gap of unknown length
*	95301	106211:	contig of 10911 bp in length
*	106212	106311:	gap of unknown length
*	106312	114409:	contig of 8098 bp in length
*	114410	114509:	gap of unknown length
*	114510	124101:	contig of 9592 bp in length
*	124102	124201:	gap of unknown length
*	124202	136041:	contig of 11840 bp in length
*	136042	136141:	gap of unknown length
*	136142	155777:	contig of 19636 bp in length
*	155778	158877:	gap of unknown length
*	158878	179225:	contig of 23348 bp in length
FEATURES			
source	Location/Qualifiers		
	1. .179225		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/clone="RP11-511P7"		
	1. .1315		
misc_feature	/note="assembly_name:Contig25"		
	1416. .2954		
misc_feature	/note="assembly_name:Contig27"		
	3055. 4541		
misc_feature	/note="assembly_name:Contig28"		
	4642. .6249		
misc_feature	/note="assembly_name:Contig29"		
	6350. .8039		
misc_feature	/note="assembly_name:Contig30"		
	8140. 9473		
misc_feature	/note="assembly_name:Contig31"		
	9574. .11903		
misc_feature	/note="assembly_name:Contig32"		
	12004. .14193		
misc_feature	/note="assembly_name:Contig33"		
	14294. .17408		
misc_feature	/note="assembly_name:Contig34"		

```
misc_feature 1. 1315
/note="assembly_name:Contig25"
misc_feature 1416. 2954
/note="assembly_name:Contig27"
misc_feature 3055. 4541
/note="assembly_name:Contig28"
misc_feature 4642. 6249
/note="assembly_name:Contig29"
misc_feature 6350. 8039
/note="assembly_name:Contig30"
misc_feature 8140. 9473
/note="assembly_name:Contig31"
misc_feature 9574. 11993
/note="assembly_name:Contig31"
misc_feature 12004. 14193
/note="assembly_name:Contig33"
misc_feature 14294. 17108
/note="assembly_name:Contig34"
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-511P7"
1..1315
/note="assembly_name:Contig25
1416..2954
/note="assembly_name:Contig27
3055..4511
/note="assembly_name:Contig28"
4642..6249
/note="assembly_name:Contig29
6350..8039
/note="assembly_name:Contig30
8140..9473
/note="assembly_name:Contig31"
9574..11903
/note="assembly_name:Contig32
12004..14193
/note="assembly_name:Contig33
14294..17408
/note="assembly_name:Contig34"

```



```
misc_feature      17509. .20169 /note="assembly_name:Contig35"  
misc_feature     20270. .22745 /note="assembly_name:Contig36"  
misc_feature     22846. .26671 /note="assembly_name:Contig37"  
misc_feature     26772. .30438 /note="assembly_name:Contig38"  
misc_feature     30539. .33209 /note="assembly_name:Contig39  
clone_end:r7  
vector_side:right"  
misc_feature     33310. .36885 /note="assembly_name:Contig40"  
misc_feature     36986. .41449 /note="assembly_name:Contig41"  
misc_feature     41350. .45860 /note="assembly_name:Contig42"  
misc_feature     45961. .49359 /note="assembly_name:Contig43"  
misc_feature     49460. .54617 /note="assembly_name:Contig44"  
misc_feature     54718. .61279 /note="assembly_name:Contig45"  
misc_feature     61380. .66606 /note="assembly_name:Contig46"  
misc_feature     66707. .73439 /note="assembly_name:Contig47"  
misc_feature     73540. .80804 /note="assembly_name:Contig48"  
misc_feature     80905. .87301 /note="assembly_name:Contig49"  
misc_feature     87402. .95200 /note="assembly_name:Contig50"  
misc_feature     95301. .106211 /note="assembly_name:Contig51  
clone_end:SP6  
vector_side:left"  
misc_feature     106312. .114409 /note="assembly_name:Contig52"  
misc_feature     114510. .124101 /note="assembly_name:Contig53"  
misc_feature     124202. .136041 /note="assembly_name:Contig54"  
misc_feature     136142. .155777 /note="assembly_name:Contig55"  
misc_feature     155878. .179225 /note="assembly_name:Contig56"  
misc_feature     179225. .20169 /note="assembly_name:Contig57"
```

Query Match	42.5%	Score 1256.4	DB: 58	Length 179225
Best Local Similarity	98.2%	Pred. No. 3.7e-185		
Matches 1344	Conservative	0	Mismatches 16	Indels 8
				Gaps 7
QY 1561	cggcgatcacacacccggagagaaagccttaagctctgcccagactgcygcaaaagccttcagc	1620		
Db 124221	CTGCCATTCACACCGCGAAGAAATCCCTACGTCTCCCGAACT--TGCCAAACCTTCAGC	124278		
QY 1621	cagaagttcaacactggtgtgcagccggcgatcacaaagggagagcgagcgccctaacgctgt	1680		
Db 124279	CAGAAGTTCACACTGTGTGTGCGACCGGCGCATTCACACGGGGAGGGGCGCTTCACGCTTCT	124338		
QY 1681	ccc_gactcygagccgcaagcttcagccagaaagtcacacctcattcacccacgcaaaagcca	1739		
Db 124339	CCCMACATCGCGACCGACCTTCAGCCAGAAAGTCACAACTCATCACCCACCGCAAGAGCCA	124398		
QY 1740	catcgaggagagcgagcctctctgtgtgcaatctgtgtggcagagacttcgaagaagaagag	1799		
Db 124339	CATCGGGAGAGCGCGCTTCTGTGTGTGCCATCTGTGTGCCAGACCTTCGACGAGAGAGAGAG	124458		
QY 1800	actcctggccaccacgaagaagcagatgtctgaagaaggtggcgggcgccgtgtgtgctgt	1859		

Dh	124459	ACTCTGGCCCAACGAAGAAAGCACGATGTTGAGCGTGTGGGCGGTGTGGCTG	124518
Oy	1860	agaagagcctgaggctccctcctgctggagtcagcagtcggcctgaggctgcctcctagt	1919
Dh	124519	AGAGAGGGCTGGGGCTCTCTCGTGGTGGGAGTGCAGTGGGCTGGGGGTGGCTCGCTA	124578
Oy	1920	ctgagctagagggaacatggaaaccctagaagggaatggaagatgcggaggatggcctgg	1979
Dh	124579	CTGGAGTGGGGGACAATGGCAATCTGTAGAGGGGATGTGAGAACCGCGGAGTGA	124638
Oy	1980	gggcctctgctgagagagagtcaccccgctggcagggaaccaccttcacagcgag	2039
Dh	124639	GGGCCCTCTAGCCAGAGAGAGGTCAACCCCGGTGGCCAGGGAAACCATCTTCAAG	124698
Oy	2040	gaagcagcagcccccagcctgctgtgctcaaggctccgctcagtcagcctgtgccttga	2099
Dh	124699	GACGGCGGCTCCAGCTGTGTGTGTCTAAGGCTCCGCTCTGACTCGCCCTGTGCTTGA	124758
Oy	2100	aagcagcaata-catccgcccctacagccctctgctctagaggccacagctggaaag	2158
Dh	124759	AAGCAGCAATGTAGCATCCGCCCTTAGAGCCCTGTGGCTAGAGGACCAAGTGA	124818
Oy	2159	aag-cccccactctctggtatcaagccttaatgccccgctctcttaactgtaagtact	2217
Dh	124819	AAGACCCCTCAATCCCTGTGTGTAAAGCCCTTAAAGCCCGTCTCTTAACTGTAA	124878
Oy	2218	t-agaatcttttggaaagcagcgctgtgtatagctctttaaataagctcttggcctag	2276
Dh	124879	TAAATCAATTTTGGAGACAGACGGCTGTGATGATCTCTTAAATGAATCTCTGG	124938
Oy	2277	acagcttggagaacctgctgtgccttgttaagacagaacttggcccttgcagagcaaga	2336
Dh	124939	ACAGCTTGGAGAACCTGCTGGCCCTTGTAAAGACACACTTGGGCTTTGCAGAGCA	124998
Oy	2337	ggtgaagcgaagccaactcttaacctctccctccccaactgccccctgctglaaggac	2396
Dh	124999	GGTGAAGGAAGCCACTTTCACCTCCCTCCCTCCCACTGCCCCCTGCGTAGAGCAC	125058
Oy	2397	ccaagcttggaggaagaccgctgtgttaataacttcacactctctctcccaagaagca	2456
Dh	125059	CCAGACTTGGAGAGACCCGCTGTGCTGTAAATCAATCCATCTCTCTTCCCAAGACA	125118
Oy	2457	gaccacaaggaattactcctctgtgtctgtctcgtttatctgtgcgccccccagcgctg	2516
Dh	125119	GATCCCAAGGCAATTAATCTTGGTGTGTGTGCTGCTTATGTGTGTGCGCCATCCAG	125178
Oy	2517	agaagctcccccctgctctcagagaacatgtgtccagagcctctgtcgaacacgcagccc	2576
Dh	125179	AGAGCTCCCTCGGTGTGCACAGACACAGTGTCCAGGCTCTGTCTTAAACCGCACCCC	125238
Oy	2577	ctccctgtcctctcccaagctagaatgttcaaggagaagatgcgcatgtgtgtagtga	2636
Dh	125239	CTCTTTCCTCTCTTCCAGAGTCCAGCATGTACGCGAAGGACATCCCATATGGATGGA	125298
Oy	2637	gggccaagctggagggaagtgtcgtgtgagtctcc-ttctccatcttctagaacat-gaca	2694
Dh	125299	GGGCGACGTGAGGGGAATGCTGTGTAGTTCCTTTTCTCCATTTCTTACATATG	125358
Oy	2695	ccctgagcctctgtgagcaattagtgacagaagaaactccgaacctctgagccctgagat	2754
Dh	125359	CCTGGCCCTCTGCTTGGAGACATTAAGGTGACAGAACTTCCGACACTCTCGAGCGCCT	125418
Oy	2755	gattctaatcttttggaaattctaattgttagaaactctctctaataatgatatctgt	2814
Dh	125419	GATTTAATTTGTGAAATTTCTAATTTGTAAATAATCCCTTATTAATGAATGAATTC	125478
Oy	2815	cttctcctaataattctacctaattggcccttctgtctctcttggaaactaaacgaacaac	2874
Dh	125479	CTTCTCTAATTTCTTCTACCTATTGTGGCGCTTGTCTGTCTCTGSAACCTAAACGA	125538
Oy	2875	catttaaccctcccttccaacactagagataaacattgtgttttagaa-2922	

Db 125539 CATTACCCCTCTTCACAACTAGACAAATTAAGATTGTTTACGAA 125586

## RESULT 6

AC015887

LOCUS AC015887 201458 bp DNA HTG 30-MAY-2000  
Mus musculus clone CT7-315E6, WORKING DRAFT SEQUENCE, 20 unordered  
pieces.

## ACCESSION

AC015887

## VERSION

AC015887.3 GI:8112888

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

house mouse.

## ORGANISM

Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

Barren,B., Linton,L., Nusbaum,C. and Lander,E.

1 (bases 1 to 201458)

Mus musculus chromosome, clone CT7-315E6

Unpublished

2 (bases 1 to 201458)

Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barina,N., Beckert,J., Boguslavsky,L., Boukhalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Doyle,M.,  
Cooke,P., DeRubeis,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehocky,J., Lieu,C., Locke,K., MacDonald,P., Margulis,N.,  
McEwan,P., McGurk,A., McKernan,K., McDonald,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessier,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 30, 2000 this sequence version replaced gi:7249370.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center Project name: L742

Center Clone name: 315\_E-6

Sequencing vector: M13; M7815; 44% of reads

Sequencing vector: Plasmid; n/a; 5% of reads

Chemistry: Dye-terminator Big Dye; 42% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 19261 bases at least Q40

Consensus quality: 196150 bases at least Q30

Consensus quality: 197810 bases at least Q20

Insert size: 182000; agarose-gel

Insert size: 195558; sum-of-coverage

Quality coverage

NOTE: This is a 'working draft' sequence. It currently

consists of 20 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1060: contig of 1060 bp in length

\* 1061 1160: gap of 100 bp

\* 1161 2600: contig of 1440 bp in length

\* 2601 2700: gap of 100 bp

## FEATURES

source

\* 2701 4058: contig of 1358 bp in length  
\* 4059 4158: gap of 100 bp  
\* 4159 5831: contig of 1673 bp in length  
\* 5832 5931: gap of 100 bp  
\* 5932 7380: contig of 1448 bp in length  
\* 7381 7480: gap of 100 bp  
\* 7481 8820: contig of 1340 bp in length  
\* 8821 8920: gap of 100 bp  
\* 8921 10519: contig of 1599 bp in length  
\* 10520 10619: gap of 100 bp  
\* 10620 13322: contig of 2703 bp in length  
\* 13323 13422: gap of 100 bp  
\* 13423 15194: contig of 1772 bp in length  
\* 15195 15294: gap of 100 bp  
\* 15295 18405: contig of 3111 bp in length  
\* 18406 18505: gap of 100 bp  
\* 18506 20293: contig of 1788 bp in length  
\* 20294 20393: gap of 100 bp  
\* 20394 24037: contig of 3644 bp in length  
\* 24038 24137: gap of 100 bp  
\* 24138 30849: contig of 6712 bp in length  
\* 30850 30949: gap of 100 bp  
\* 30950 44871: contig of 13922 bp in length  
\* 44872 44971: gap of 100 bp  
\* 44972 61254: contig of 16283 bp in length  
\* 61255 61354: gap of 100 bp  
\* 61355 74139: contig of 12785 bp in length  
\* 74140 74239: gap of 100 bp  
\* 74240 99150: contig of 24911 bp in length  
\* 99151 99250: gap of 100 bp  
\* 99251 129995: contig of 30745 bp in length  
\* 129996 163192: contig of 33097 bp in length  
\* 163193 163292: gap of 100 bp  
\* 163293 201458: contig of 38166 bp in length.

Location/Qualifiers

1. .201458

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_id="CT7-315E6"

/clone\_lib="CTB Mouse BAC"

1. .1060

/note="assembly\_fragment"

1161. .2600

/note="assembly\_fragment"

2701. .4058

/note="assembly\_fragment"

4159. .5831

/note="assembly\_fragment"

5932. .7380

/note="assembly\_fragment"

7481. .8820

/note="assembly\_fragment"

8921. .10519

/note="assembly\_fragment"

10620. .13322

/note="assembly\_fragment"

13423. .15194

/note="assembly\_fragment"

15295. .18405

/note="assembly\_fragment"

18506. .20293

/note="assembly\_fragment"

vector\_side:left"

20394. .24037

/note="assembly\_fragment"

24138. .30849

/note="assembly\_fragment"

30950. .44871

/note="assembly\_fragment"

44972. .61254

/note="assembly\_fragment"





[illegible][illegible]

|||||

Db 279 GTGCTGAGAGAGGAGCAATGGAATCTAGAGAGGATGAGAAAGCGGGAGAGTGAAGCTG 220

Qy 1977 ggtggagccctgtcagagagagaggtcaaccccggtgagccaggaacccacttccaaagcgc 2036

Db 219 GGTGGGCGCTGTCTAGAGAGAGAGGTCAACCCCGGTGGCCAGGAGACCCACTTCCAAAGCGC 160

Qy 2037 aggaagccgagccctcagctgtgtgtgtcctaagcctccgctcagctcagctgtgcccgt 2096

Db 159 AGGAGAGCGCGGCTCCAGCTGGTGTGTGCTTAAGGCTCCGCTCGACTGCCCTGTGCCCTG 100

Qy 2097 gaaaagcaagcaata.caccgcgccttaccagccctgtgctagaggaagccaccagctggaa 2155

Db 99 GAAAGACAGCAATAGATCGCGCCCTTAGAGCCCTGCTGAGAGAGAGCACCAGCTGGAA 40

Qy 2156 aggaag-ccctcagctcctgtgtattaaagccctaatgc 2193

Db 39 AGGAGAGCGCTCCAGCTGTGCTGTAAAGCGCTTAATAC 1

RESULT 8

LOCUS AB002324 6045 bp mRNA PRI 13-FEB-1999

DEFINITION Human mRNA for KIAA0326 gene, partial cds.

ACCESSION AB002324

KEYWORDS AB002324.1 GI:2224592

SOURCE KIAA0326.

ORGANISM Homo sapiens male brain cDNA to mRNA, clone\_11b:pbjunescriptII SK plus clone:H0579.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

Location/Qualifiers

1.6045

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="HG0579"

/clone\_11b="pbjunescriptII SK plus"

/sex="male"

/tissue\_type="brain"

1.2784

/gene="KIAA0326"

<1.2784

/gene="KIAA0326"

/codon\_start=1

/protein\_id="BAA20784.1"

/db\_xref="GI:2224592"

/translation="RVLSLESPLEKNDPRVLAQSVPRGALKLSPGLGDSARFLRPD  
PRAGPWNTAVLSGMEPEPALWPDLOGEOPNDHNRGAESSENEESRPSGSEI  
IGDPAESPESSDSEMSRSDSPVQNPPTPLGHSNPLDHOIPLDPAEVPYT  
PDDWTKAASWOMGALTWNBPVPAPEPSIRELVGRRGAGAEKPYICNCGKSES  
QMSKILRHRIHTGERPNTSCGKSFYSSSHVQHORHTGEEKPKCDCKCEFS  
SNLVOHORHTGEEKPKCEKAEAFYOSNLIKHOSHTGEEKPKCEKRAVRS  
LHOATHTGEEKPKCEKRRGONHNLKHKHAGKPRVCEKSGFSYSSLS  
OHOHTHTGEEKPKCEKSGFSYSSLSLHQRTHLEDFKPCVCGKTYLSTLIRH  
ORHTHTGERPKCEKSGFSYSSSLNHQIRHGERPYICADCGKSFINSSTLIRH

HTGEEKPKCDCKSGFSIRSSHLIQRHTHTGEEKPKCEKCEKSFSSSLNLTHTVHTH  
MDELVFCSDCKAFLEAEHLEBHRVTHGERKTPARAGDSILGIGDSILTPREGA  
KPKKICGCKGPNDEGIFMOHORHTHTGENDPRADGLTAAAPKPOLSPRLPREGN  
STPGAEBGAERAPGDLKPRBEGESQORGLISSTYICSHCESEFLDRSVLLOHOL  
THNEKPELPEPDRIGLEGAGSPLEKPCPKCPSKGLISLHLHOKVHAGK  
SOKSPBELKSSSVLLEHLHRSPLGARPKSDRAFLRVALTRHOETHTOKPKNP  
EDPEPAVLTSTDOEGEGEPTPTPESSSHGEGNPTLVEEKPYLCEKAGTTEVAA  
LILHRSCHGVSL"

BASE COUNT 1514 a 1775 c 1684 g 1072 t

ORIGIN

Query Match 10.7%; Score 315.8; DB 8; Length 6045;

Best Local Similarity 55.8%; Pred. No. 1,1e-39;

Matches 668; Conservative 0; Mismatches 517; Indels 12; Gaps 3;

Qy 633 cccctccgccccgagagcccgccctcatalatgcggaactgtgcccggagcttgcga 692

Db 597 CCGCCGCGCGGGGGGAGAAAGCCCTACATCTGCAAGAGATGGCGGCAAGAGCTTCAAGCA 656

Qy 693 gtggagccagctagttgcccacaagcggtgcagctgtaggcccctgtaggaagcgccg 752

Db 657 GTGCTCAACCTCTGCGGACCAAGCGCATCACAGGAGAGCGGCCCAACACTGCTC 716

Qy 753 agccaagctctggtggccccgccccgagggcccccgcccggtgtaccgcccccgccccg 812

Db 717 CGAGTGGCGGCAAGAGCTTACAGAGAGCTGCACTGTGTCAAGCACCAGCGGACGC--- 772

Qy 813 tggagatgctgcagcccccctcagctgtgtgtgtgtggaagcagcttccgcgcaaa 872

Db 773 -----ACACCGGGAGAGAGCCCTACAAAGTGCCCGAGTGGGGAAGTCTTCAAGCTGAG 827

Qy 873 gcccaactgtatgcgttaacgcgcggtgcacaagcgccgagccgcccccaagctgccccga 932

Db 828 CTCCAACCTGTGACAGCAGGACGACGACGAGGAGAGAAAGCCCTTAAGTGCACAGA 887

Qy 933 gtgcggaagcgcttaccataaagccctcagctgtgtgtgtgtggaagcagcttccgcgca 992

Db 888 GTGCGAGAAAGCTTCAACAGAGCAGACCACTTCATATAGCACCAGAGATCCCAACACCGG 947

Qy 993 cgaagaagcccttaccggtgcgaagagtggtgcgcgcgcttccgcgacaaacccaactgct 1052

Db 948 CGAAGAACCTTCAAGAGGCGGCGAGTGCCGCGGCTTTTACCGGAGCTCGGACCTCAT 1007

Qy 1053 gtctcaagaagaatltacaagagatccgaggggtgtgccccagcgcccccgccccgg 1112

Db 1008 CCAAGCAGCAGGACCAAGCCACAGCGGAGAAAGCCCTTAAGTGCAGTGCAGGAGAGCG 1067

Qy 1113 gaggccccagctgcagcgcccccgagaggttcgcggcgccgagcccaaccccg--ggggt 1170

Db 1068 CTTGCGCCAGAACCAACCACTCTCAAGACACAGAGATCCACGCGGCGAGAGACCATATA 1127

Qy 1171 cctctgaagccgccccagagcgccgccccgagggcccccgcaagagaccc--gaagagccc 1229

Db 1128 CCGGTGACAGGAGTGCGGGAAGAGCTTCAACAGAGCTGCGAGGTGACCCAGCAGCAGCG 1187

Qy 1230 gacgaagccccccctcctcctacagctgcagagctgcggaagagcttccggtctga 1289

Db 1188 CAGGACACAGAGGAGAGAGCCCTAAGATGCTTGAAGTGGGCAAGAGCTTGGCCACAG 1247

Qy 1290 ggcgttcctgcggccccagcgagcgcaacacggggaagcgccctcagctgcgscga 1349

Db 1248 CTCACACCTTATCAAGACACAGGAGCTACCTGCGGAGAGACCCGTTCAAGTGCACAGT 1307

Qy 1350 gtgcggaagaacttgcggaagaagaagcaacactgtgtgcgcaactgcggttgaactcgg 1409

Db 1308 GTGGGGAAGAAGCTTCAAGCTGAGCGGACAGCTTGTCTGCGGACACAGGACGACAGCGG 1367

Qy 1410 cgaagcgcccttgcgctgcggaagagtggtgcgcgcgcttccccaaggaagcaactgcg 1469

Db 1368 CGAGCGGCGCTTCAAGAGTGCAGAGAGCTTCAAGGTCTAGCTTCAACCTCAT 1427

Qy 1470 ggcgcatcgccggaacgaagcccccgatcgcccttgcgtgtgtccgacgtgcggaagcg 1529

Db 1428 CAACACACGAGCATCCACCGCGAGCGGCCCTACATCTGCGCGAGCTGCGCAAG 1487  
1530 cttccgacacaaacctccctcggcgacccgacatccacacggcgagaaacctta 1589  
1488 cttatcatgagctccaccccttattccgacacgacatccacacggcgagaaacctta 1547  
1590 cgtctgcccagctcgcgaagccttcagcgaagtcacacacctgtgtcgcacggcg 1649  
1548 CAAGTGTCCGACATCGCGGCAAGAGCTTCATCCGACGCTCCACCTTATCCAGACCGCG 1607  
1550 catcacacggcgagcgagcgacctcgtccgactgcgacggcgacggcgaagaa 1709  
1608 CACGCACACCGCGGAGAGCCCTACAGTCCCGCGAGTGGCGGACAGCTTCAGCGCAG 1667  
1710 gtcaacctatcacccacggcgaagacacatccggaacggcgacctctgtgtgcat 1769  
1668 CTCACACCTTATTCACCGCGCGACGACGACGACGACGACGACGACGACGACGACG 1727  
1770 ctgtgacgacacgtcgcgacgagagagactcctgacccagagaagacagca 1826  
1728 CTGGGGAAGGACCTTCCCTGGAAGCCACGACGACGACGACGACGACGACGACGACG 1784

RESULT 9  
AC013570/c  
LOCUS  
DEFINITION Homo sapiens clone Rpl1-2C24, WORKING DRAFT SEQUENCE, 37 unordered  
pieces.  
AC013570  
VERSION AC013570.4 GI:8072509  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 185306)  
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baldwin, J., Barne, N., Beckler, R., Boguslavsky, L., Bouknight, B.,  
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., Dekrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
Ferreira, P., Fitzhugh, K., Forrest, C., Funke, R., Gage, D., Galagan, J.,  
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J., Johnson, R., Jones, C., Kahn, L., Karulis, A., Klein, J.,  
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwan, P., McGurk, A., McKernan, K., McDonald, J., Meidrim, J.,  
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tefaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced g1:7654731.  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W1BR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L2523  
Center clone name: 2\_C\_24  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
Consensus quality: 162201 bases at least Q40  
Consensus quality: 173731 bases at least Q30  
Consensus quality: 178289 bases at least Q20  
Insert size: 196000; agarose-fp  
Insert size: 181706; sum-of-contigs  
Quality coverage: 3.6 in Q20 bases; agarose-fp  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs  
-----  
NOTE: This is a 'working draft' sequence. It currently  
consists of 37 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
1 1008: contig of 1008 bp in length  
1009 1108: gap of 100 bp  
1109 2471: contig of 1363 bp in length  
2472 2571: gap of 100 bp  
2572 4482: contig of 1911 bp in length  
4483 4582: gap of 100 bp  
4583 6444: contig of 1862 bp in length  
6445 6544: gap of 100 bp  
6545 8758: contig of 2214 bp in length  
8759 8858: gap of 100 bp  
8859 10776: contig of 1918 bp in length  
10777 10876: gap of 100 bp  
10877 12177: contig of 1301 bp in length  
12178 12277: gap of 100 bp  
12278 13361: contig of 1084 bp in length  
13362 13461: gap of 100 bp  
13462 15267: contig of 1806 bp in length  
15268 15367: gap of 100 bp  
15368 16785: contig of 1418 bp in length  
16786 16885: gap of 100 bp  
16886 18775: contig of 1890 bp in length  
18776 18875: gap of 100 bp  
18876 21003: contig of 2128 bp in length  
21004 21103: gap of 100 bp  
21104 22763: contig of 1660 bp in length  
22764 22863: gap of 100 bp  
22864 25665: contig of 2802 bp in length  
25666 25765: gap of 100 bp  
25766 28291: contig of 2626 bp in length  
28292 28391: gap of 100 bp  
28392 29739: contig of 1348 bp in length  
29740 29839: gap of 100 bp  
29840 33467: contig of 3628 bp in length  
33468 33567: gap of 100 bp  
33568 37139: contig of 3572 bp in length  
37140 37239: gap of 100 bp  
37240 41052: contig of 3813 bp in length  
41053 41152: gap of 100 bp  
41153 45504: contig of 4352 bp in length  
45505 45604: gap of 100 bp  
45605 50022: contig of 4418 bp in length  
50023 50122: gap of 100 bp  
50123 54849: contig of 4727 bp in length  
54850 54949: gap of 100 bp  
54950 59045: contig of 4096 bp in length  
59046 59145: gap of 100 bp  
59146 65075: contig of 5530 bp in length  
65076 65175: gap of 100 bp  
65176 71663: contig of 6488 bp in length  
71664 77852: contig of 6089 bp in length  
77853 77952: gap of 100 bp  
77953 83219: contig of 5267 bp in length  
83220 83319: gap of 100 bp  
83320 88571: contig of 5252 bp in length  
88572 88671: gap of 100 bp



Accession	Sequence	Position
D80785	CTTCATCATGAGCTCCACCCCTTATCCGCGACACAGGAGCTCCACACCGGAGGAAACCCCTA	80726
OY 1590	cgctgcgccgcgacthcgggcaaaagccttcagccagaagctccaaactcgtgtgcacagcg	1649
D80725	CAAGTGTCCACACTCCGGCAAGAGCTTATCCGACGCTCCACCTTATTCACACACCGCG	80666
OY 1650	catccacacgggcgagcgagccctacgcctctgcgactctgcgactcgacgcgcaacttaagccagaa	1709
D80665	CACGCACACCGCGGAGAGCCCTTACAACTGTGCCCGGCGGCGGCAAGAGCTTCACGCCACAG	80606
OY 1710	gtccaacctcatcaccacccgaagagcagcatcccggaagcgagcctctgtctgtgcacat	1769
D80605	CTCCAACTTATTACCAACGATCCGACGACACATGAGACGAGAAACGTTCGTGTGCTCCGA	80546
OY 1770	ctgtggtgcagagacttcagacgagcagagagaagactcctctgtgccacacagaagaagacaga	1826
D80545	CTGCGGGAAGGCTTCTTCGGAAGCCACACAGCTGAGGACGACCGGCGTATCTCAATGA	80489

SULT 10  
SAJ03147/c

LOCUS	HSAMJ03147	239566 bp	DNA	PRI	11-DEC-1998
DEFINITION	Homo sapiens complete genomic sequence between DISC3070 and DIS62275, containing Familial Mediterranean Fever gene disease.				
ACCESSION	AF003147				
VERSION	AF003147.1	GI:2808656			
KEYWORDS	HOMN4 gene; mareno gene; marenostrin; metalloproteinase; mmp20				

```

/translati0n="MPPGVTATMKRPRCSLPDVLGVALVRRRRYALSSGVKKRPL
TWRRSPGSSQJSGTFRWNPSTALMAMGMSGGLTFHEVSPGQEDDILDFARAF
HQDSYPDGJGLTAHAFRPGSLPIGSDTHFDEETWTFGSKASQJLEGLAGSSPYD
EELGSGRGWRVNDLPGSPBRUS"
6776. 9221

```

```

source
1. .36906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="p13.3"
/clone="YAC 16g7"
/sub_clone="bc2"
2959. .3265
/standard_name="HSA353YH1"
6759. .6775
/exon
/gene="mmp20"
/number=1
join(6759. .6775,9222. .9357,9467. .9758,15042. .16530)
/gene="mmp20"
join(6759. .6775,9222. .9357,9467. .9758,15042. .16530)
9467. .9758,9759. .15041,15042. .16530)
/gene="mmp20"
join(6772. .6775,9222. .9357,9467. .9758,15042. .15161)
/gene="mmp20"
/codon_start=1
/product="metalloproteinase"
/protein_id="CA05902.1"
/db_xref="GI:2808657"
/db_xref="SPTREMBL:O43923"
CDS

```

intron	9759..15041 /gene="tmmp20" /number=3
mRNA	complement(14113..16533) /gene="tmmp20" /note="transcriptional u expression only detectab not from liver}" complement(14113..16533)
gene	/gene="tmmp20" 15042..16530 /gene="tmmp20"
exon	/number=4 25020..28566 /gene="HUNKK4" 25020..28566 /gene="HUNKK4" complement(26351..26612)
STS	/gene="HUNKK4" standard_name="HSB070YG complement(26351..26612) /gene="HUNKK4"
gene	31077..67877 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="P13..3" /clone="YAC 16q7" /sub_clone="bd6"
source	46940..23956 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="P13..3" /clone="YAC 633d12" /sub_clone="aa6"
variation	53459 /note="aa6" /replace="a"
variation	53486 /note="aa6" /replace="c"
variation	57934 /note="aac" /replace="g"
variation	58731 /note="aa6" /replace="c"
variation	61447 /note="aac" /replace="c"
variation	62401..62404 /note="aac" /replace="ac" 66755..108346 /organism="Homo sapiens" /db_xref="taxon:9606"
source	

**Source**

	/chromosome="16"
	/map="p13.3"
	/clone="YAC 26fe7"
	/sub_clone="30c3"
	90616..137018
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="p13.3"
	/clone="YAC 26fe7"
	/sub_clone="30e1"
	126806..170318
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="p13.3"
	/clone="YAC 26fe7"
	/sub_clone="30g4"
	complement(161861..161970)
	/standard_name="CHIC.A19A1E08.P34246"
	163476..164414
	/gene="hsolfmf"
	163476..164414
	/gene="hsolfmf"
	/codon_start=1
	/number=1
	/product="Olfactory receptor"
	/protein_id="CAA05903.1"
	/db_xref="GI:2808658"
	/db_xref="SPTRMBL:043749"
	/translation="MSGTQSSVSSEFLLGLSRPOOHLLFFELSMYLATVGNLL ILTSIDISCLHTPMYEFPLNSLFVDICEFTTPVKMLAHNILETQTISEGCTLOYW EVFMVEDADNLALVMAVDHFVAACHPLHYAKMTHQCALVAGLVANLNLHTT LIMALPSCDNNAILTFHCDDVTPLIKLSCSDPHNEVTLISEGALVMITPELCIASY NHITQTVKVFSTGKRKAFFSTGCSGHNVLLFYSTIAAYFNPLSSHSAEKDMATIV LYTVTPEMLNPFIYSLSRNRYLKALKAKVYGAVESV"
	173301..217867
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="p13.3"
	/clone="YAC 26fe7"
	/sub_clone="30f9"
	174780..175778
	/gene="Olfmf2"
	174780..175778
	/gene="Olfmf2"
	/codon_start=1
	/pseudo
	/number=1
	/product="Olfactory receptor pseudogene"
	/db_xref="PID:e1310000"
	179596..222837
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="p13.3"
	/clone="YAC 26fe7"
	/sub_clone="30e10"
	complement(183121..194589)
	/gene="znfmf"
	complement(join(183121..183842,191650..191773, 192062..192150,192735..192984))
	/gene="znfmf"
	/codon_start=1
	/product="zinc finger protein"
	/protein_id="CAA05905.1"
	/db_xref="GI:2808659"
	/db_xref="SPTRMBL:015361"
	/translation="MMAAKVMPMPKPQSFILAVPPDSKLGGDLLDATNGPTTHO LVLEHFLIFLPRLPSIVQSOKYTELVIMRKVVSSLQRVRPRPLVKILPRGVAKDEE VLVLKNAPELVVEFDLNVFHQCEVSLDPQTQLTSEKEDSSVGEMLLAVNGSN PGSEEPREPELVNEYRKSDDDEMDSLVSQQPDNQEKERLNTSIPQKKRRNLL

Query Match		10.7%;	Score 315.4;	DB 77;	Length 239566;
Best Local Similarity		58.5%;	Pred. No. 5.2e-40;		
Matches 616;		Conservative	0;	Mismatches 416;	Indels 21; Gaps 3;
QY	794	tgacgcccccccgccggtgtagatgtccgtgcagcccccttcagtgtgctgtgtg	853		
DB	49558	TSAAAGCGGACACAGCTGGCGTGGCACCGGCGGGCGGCTCTGTCGCTTCTGCG	49449		
QY	854	gcaagcgcttcccgcgcaaacgcccattgattcgtctacccgcggtgcacagggcgagc	913		
DB	49498	GGAAAGCTTTCGCGCGCGAGCTCCATTTCACAGCTGCACATGGCGACTACAGGCGAGC	49439		
QY	914	ggccccacacagtgcccgccgagttgcggaagcggtttaccatcgaagccctatcgaattcgc	973		
DB	49438	GGCCGCGACCGCTTCGACCTGTGCGGCGCACCGCTTCCGCGAGCTGCGCACTGACCAAGC	49379		
QY	974	accggtgcataccacacgscgsgaagaccctacccgtgcacaaagtgtcgscgcgcttcc	1033		
DB	49378	ACCTGCTGACCCACTCTCTCCGAACCCGGCTTCTGTGGCGGAGAGTGGCGGCGGCTTCC	49319		
QY	1034	ggcacaaccca-----actgtgtctcacaagaagattaca-----	1073		
DB	49318	AGCGCGCGCGCGCTTGTGAGCAGCACTGCTGCGCGCACCGCCAGGACCAAGACCGCCT	49259		
QY	1074	ggatccgaggggtgtgcgccagggccgccccggccccggggagcccccaagcttgcagccg	1133		
DB	49258	GGGCTCTGAGAGTAAGCGCGAAGCGCGCGCACCTGACCGATGCTGTGCTCCACTGCG	49199		
QY	1134	cccccaagatgtcggcgccgagccccaccgcggtgtactctgaaacgscgccagagcc	1193		
DB	49198	GCCAGAGCTTCCAGGCGCGCTCCAGCTTCAGAGCGGCACCTGGGATCCAGCGGAGGACA	49139		
QY	1194	ggcgccaggggcccccgcaagagaccgcgaaagaccg--atogaagccccccctctct	1252		
DB	49138	AGGACCGCGCGGTCTCTCCGAAGGCTCCGCGACCGCGCGGAGACTCCAGCGAGGCGCT	49079		
QY	1253	acaagttgacagcactgcgsgagagagattcccgcttgcgttgagcgcttcctgsgggccacagc	1312		
DB	49078	TGCTGTGCACGCACTGCGCAAGGCTTCGCGCGCAGCAGACCACTGTGTCGCCACCGGA	49019		
QY	1313	ggcagacacccgggagcgcccttcaactgtgcgagtggtgcggaagaagattcgcgcaaga	1372		
DB	49018	GGGTGCACACGGGGCAGCGCGCTTCTCTCTGCGCAAGCTTGGCGCGCAGTTCAAGCAGA	48959		
QY	1373	agacgaacactgtgtgcgaactcgcgcgtgtgacttccgsgcgagcgcccttgcgttgcgag	1432		
DB	48958	GCTTCAGCTGTGTCCACCAACAGGGGTGCAACAGGGGAGAGAGCCCTTAAGCTTCTCCG	48899		
QY	1433	atgtgcgcgcgcgtcttcccaaggcagcatctgtggcggtgcattcgcgcgacccagcc	1492		
DB	48898	AGTGGGGGAAGGGCTTGTGTGCGCCGGGCGAGCCTTGGCCGCACTGTGTGACCCAGGTG	48839		
QY	1493	cgatgcgccttctgtgtgtcccgactgcggaagggcccttccgcgaacaaacctcactcgg	1552		
DB	48838	GGCCTTGCGGCCACCACTGCACCCAGTGGCGGGAAGAAGTTTCGCGCAAGCCAGATCTGG	48779		
QY	1553	cgcgacacccggtcgtatccacacacgsgcgaagaccctacgttctgtcccgactgtgcgcaag	1612		
DB	48778	CCCGCACACGAGGCAAGCCACACGAGGGCGAGAAACCCCTTGGCGGTGCAAGTGTGAGG	48719		
QY	1613	cccttaagccagaagtccaaactcgtgtgtgcacccggtgcattccacagggcgagcgccct	1672		
DB	48718	GCTTACGCCAGAGCCGCCACACTGTGGGCGCGACACAGGATCCACACAGGGGAGAGCCCC	48659		
QY	1673	agcgctgtccgaactgcgacagcgagcttaccgacgaagaattcgaactatgaaccgcgca	1732		
DB	48658	AGCGCTTGCACACTGTCCGCGCACCGCTTTCGGAATAGTCCAAACTGGCGCGCCCATTCGC	48599		
QY	1733	agagccacactccgggagcgcccttctgtgcacatctgtgtgccagacttgcagacg	1792		



```

Db 48598 GCAGCCACACGGGCGGCGCCCTACGCTGACAGCGGCGGCTGCGACCTTCCGGCGCA 48539
OY 1793 aggaagactcctggccaccagaagaagacagc 1825
Db 48538 ACGCCATCTGCGGCGGCGACCTGCTACCCATG 48506

RESULT 11
AK023017
LOCUS AK023017 2382 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens CDNA FLJ12955 fis, clone NT2RP2005496, moderately
similar to ZINC FINGER PROTEIN 135.
ACCESSION AK023017
VERSION AK023017.1 GI:10434737
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA,
clone.lib:NT2RP2 clone:NT2RP2005496.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (sites)
AUTHORS Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Mishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,
Masuhu,Y. and Kanehori,K.
NEBO human cDNA sequencing project
JOURNAL Unpublished (2000)
TITLE 2 (bases 1 to 2382)
REFERENCE Isegai,T. and Otsuki,T.
AUTHORS Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isegai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction; 5'-3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source
1.2382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RP2005496"
/clone.lib="NT2RP2"
/note="cloning vector: PME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
BASE COUNT 487 a 789 c 656 g 450 t
ORIGIN

Query Match 10.7%; Score 315.2; DB 37; Length 2382;
Best Local Similarity 56.5%; Pred. No.1.8e-39;
Matches 606; Conservative 0; Mismatches 463; Indels 3; Gaps 1;

```

```

OY 936 cgggaagcgtcttacaataaagccctatctgacccgagcgcatccaccgcga 995
Db 268 CGGCAAGGCGCTTACAGCAGAGAGGTGCACCTGACCGACGAGCGCACAGCGGGCA 327
OY 996 gaagccctacccgtgcaaaagatgcygcccgttcctcgagcaaaacccaactctc 1055
Db 328 GAAGCCCTACGCGCTCCACGAGTGGCGGCAAGGCTTACAGCGAGGCGTGTACTGGCCTC 387
OY 1056 tcacagcaaatctccaaagcgtccgaggggtccgagcccgagcccgcccgaggag 1115
Db 388 CCAGTGGCGCACGACACAGCGGCGAGAGAGCGCACCGCTGCGGAGTGGGAAAGGCTT 447
OY 1116 ccccaagctgccaagccgagccccaagaa--gtccgagccgagcccaacccgaggtacc 1172
Db 448 CACGGCGCTACAGCAGCACTGACCCACACCGGCGGTGACACAGGGCGAGGGCGCTTACGC 507
OY 1173 tctgaaccgagcccaagagcgcgccgaggggcccgcgaagagaccgcaagaccgat 1232
Db 508 GTGCGCCAGTGGCGCAAGGCGCTTCCGCAACCGCTGCTCCGTATAGAGACACAGCGCAT 567
OY 1233 cgaagccccccctccctctacagctgagacgtgagcgagagcttcgcgctgagcg 1292
Db 568 CCACACCGGTGAGAGAGCCCTTACGAGTGTCCGCGCGCAAGGCTTCCGCTTCTCCTC 627
OY 1293 ctctctgagggcccaacacagcgagcgagcgagcgagcgagcttcacactgagcgagtg 1352
Db 628 AGCGCTCATCCGCCACACAGCGGATCCACAGSAGAGAGAGAGCCCTTACCGCTGGGCGAGTG 687
OY 1353 cgggaagaacttcgcaagaagaacgacactgtgtgagcaactcgcggtgacactccgcga 1412
Db 688 CGCCAAAGGCGCTTGGCGGAGATGCGCCACTGTACCCAGCAGCGGCGCGTGCACACGCGGGA 747
OY 1413 ggggccttgcttcgagagtgagtgagcgccgcttcgccagggcagacatctggcgc 1472
Db 748 GAAGCCTTACAGTGCACGAGACTGCGGCGGCGCTTTCAGCCAGAGCGGCTCTCTGCGCGA 807
OY 1473 gcatgcgcgagccacgcgcccgccttcgctgtgtccgagcttcgagcaagcctt 1532
Db 808 GCAGCGGCGGATCCACACAGCGGAGAGAGCCCTTACCGCTGCGGCGCGTGCAGAGCGCTT 867
OY 1533 ccgcaacaacccctactgagcgagcgacgcgacatccacacggcgaggaagccctacgt 1592
Db 868 CACCCAGGTGTGCGACCTGACGACACACGCGCACGCGAGCGGCGGCGCTTACCC 927
OY 1593 ctgcccgcagtcgagcaagccctttagccagaaagctcaacctgagtgacacggcgcat 1652
Db 928 CTGCCACGACTGCGGCAAGCGCTTACGCAACCGCTCCACCTCTCCAGACACCGGCTGT 987
OY 1653 ccacacggcgagcgagccttaccgtgtccgactgcgacgcgagcttcagcagaagtc 1712
Db 988 GCACACCGGTGAGGCGGCGCTTACAGGTGCTGCAATGCGGGCGCGCTTACGACGAGTGTC 1047
OY 1713 caactcatcaccacacgcgaagaagccacatccgggaagcgagcctctctgtgtgcatctg 1772
Db 1048 CTGCTTATAGAGCACAGAAAGATCCACACCGGAGAGGGCGCTTCAAGTGGGGGAGTG 1107
OY 1773 tggcgaagccttcgacgagcgagagagactcctgtgcccacgaagaagacac 1824
Db 1108 CGGCAAGCTTTCAGCCAGGCGTCTCTGCTGCGCTGCAACGCGCAGCGCAC 1159

RESULT 12
AC007228/c
LOCUS AC007228 130067 bp DNA PRI 06-APR-1999
DEFINITION Homo sapiens chromosome 19, BAC 37295 (CIT-B-21A4), complete
sequence.
ACCESSION AC007228
VERSION AC007228.1 GI:4567177
KEYWORDS HUG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



REFERENCE	Mammalia: Euthera/Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Sakaladis,G., Burkhardt-Schultz,K., Gordon,L., Dias,J., Scott,D., Stilwagen,S., Phan,H., Velasco,N., Do,L., Regala,M., Terry,R., Dangnan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J., Attix,C., Andreise,T., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B., Arellano,A., Sanders,C., Ow.D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carraro,A.V.
TITLE	Sequence analysis of a 3.6 Mb region in 19q13.4 between D19S891 to ZNF134
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 130067)
AUTHORS	Lamerdin,J.E.
TITLE	Direct Submission
JOURNAL	Submitted (06-APR-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT	Map and sequence oriented from q centromere to telomere. BAC 37295 (CIT-B-21A4) overlaps cosmid R3165 (AC003498) to the left from bases 1 to 10,159 of this accession, and is separated from BAC 330812 (CIT-B-470F8; AC006115) to the right by a sequence gap of unknown size. Additional chromosome 19 map and sequence information are available at: http://www.bio.lnl.gov/dbry/genome/genome.html.
FEATURES	Location/Qualifiers
Source	1..130067 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CIT-B-21A4" /chromosome="19" /map="19q13.4 between D19S891 and ZNF134" /clone_id="CIT-B Cal Tech BAC library" /note="LNL Clone Name: BC37295" complement(4..320) /rpt_family="MLT1B" join((332..509,682..777,1833..1996,2942..3056,3385..3480,6894..8266) /note="Hypothetical ZNF-like human protein" /codon_start=1 /evidence=not_experimental /product="R31655.2 [AA 1 - 673 ]" /protein_id="AADJ3606.1" /db_xref="GI:4567178"
repeat_region	/translation="MGMHEWLSKNVYISFGVLTFGDVAVDSEQEMEWLNPIORNI, YRKVMLENRLASLGICVSPDIVISSLEOGKEPVYRKMTKRAMCPDLKAWVKIKEL PLKKDFEGKTSOAVITERLTSLNLEYSLGEHDYALAFETGGLTKMLADVDEFO OLHAKNOCKNKIEMNSDLSGHCVAKPDIYSLLIOEKEPAMVRELTLGSFSST YAGKTRKCNCKKTFTFOSSSLTVHORIHTEGKYKCKNECGKASDDSSRRHORCHT GKRPECEIECGKAFTQNTSLIRHWYVHTGKEPKDCIDCGAFSDHIGLONORHTHG EKPCKCVCHRSFRYGSSLIVHORIHTEGKEYEDVCRAKSFSSASHASTLOHRYSGEK PFKCKECGAFFRONILHSHLRHTGKEPCACGKSFSSISOATHORHTEGKEY ECKKSAFTOKAHLOKHTEGKEYECKECCAKASOTTHILOHORYHTEGKEYPC MECKAKGDNSSCTOHORLHTGCRPYECIECGKAFKRSLSICRHSHTGKEYPCY CGKAFSRQSLSVHORLHSGKRPYECKECTFTIQIGHLNOHKRVHTGENSYNFKSR KVFEQTALAHORIHTEGESCTPSTSNPDLPFKFLMPSPSLPS"
CDS	repeat_region 2271..2562 /rpt_family="AluYo" complement(4953..4996) /rpt_family="AT_rich" repeat_region complement(5673..5793) /rpt_family="(GAAA)n" repeat_region complement(5799..5866) /rpt_family="LIMBS" 5864..5996 /rpt_family="LIMA9" complement(6099..6414) /rpt_family="AluBo" 8415..8489 repeat_region /rpt_family="MIR" complement(8977..9293) /rpt_family="MER5B" complement(9302..9497) /rpt_family="MER5A"

repeat_region	complement(9792..10090) /rpt_family="Alusg1" repeat_region complement(10096..10278) /rpt_family="AluBo" repeat_region complement(12206..13159) /rpt_family="LrR5" repeat_region complement(13352..13406) /rpt_family="(GGCA)n" repeat_region complement(13407..13697) /rpt_family="Aluub" repeat_region complement(15043..15575) /rpt_family="MLT1F" repeat_region complement(16011..16143) /rpt_family="LlPA4" repeat_region complement(16194..16338) /rpt_family="MLT1-INTERNAL" repeat_region complement(16573..16870) /rpt_family="MLT1-INTERNAL" misc_feature complement(17315..17488) /note="BLASTN similarity to G44255 (13..186); match: 0.99 score: 1.3e-63; database searched: month.na; WIAF-3903-ST5 Human THUDSON SANGER Homo sapiens ST5 genomic, sequence tagged site [Homo sapiens]" repeat_region complement(17491..17766) /rpt_family="AluBo" repeat_region complement(17772..18156) /rpt_family="LlMA4" repeat_region complement(18162..18925) /rpt_family="LlMA4" repeat_region complement(19178..19553) /rpt_family="MLT1A2" repeat_region 19599..19622 /rpt_family="AT_rich" complement(19623..19726) /rpt_family="LlINR2" repeat_region complement(20008..20389) /note="predicted exon, program: grail2exons_human_1.3, frame: 2, quality: good, score: 57.000" complement(20721..20768) /note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: excellent, score: 100.000" misc_feature complement(21509..21720) repeat_region /rpt_family="MIR" complement(21729..21817) repeat_region /rpt_family="MIR" 21897..21996 /rpt_family="MIR" 22540..22611 /rpt_family="AluBo" 22671..22874 repeat_region /rpt_family="AluBo" join((23186..23245,27268..27607,30057..31456) /note="Hypothetical ZNF-like human protein" /codon_start=1 /evidence=not_experimental /product="BC37295.1" /protein_id="AADJ3607.1" /db_xref="GI:4567179" /translation="MSQEVEVAGIKLCAMSLSGVTFPTDVADIDSODEMWLNLAQ LSLYKVMLENRLVSVGSIIFLLPHDSVFYAIIIRISHFPPOHROPLVYLFIAG LCISKDPVILILEOEKDPVYIKGMNGSLDGSKRPYCIECGAFQSALAOHORHL TGKPEPECTCGKAFFSONAHLOYHORVHTGKEYPOCKCKAKFSOLAHLOHORVHTG EKPECEIECGKAFSDSSLAHHRIHRTGKEYPCIDCGKAFSOMASLIIRRRYYHGE KPPDCIDCGKAFTDHIGLLOKRTHTGERPYKKNVCGKAFSSHSSSLTVHORHTGKERP YECNICGARSHGSLTLHORVHTGKEYPCIECGKAFROSSTLHAHORHTHTGKERP CKECSKTFESONAHLOKHTEGKEYECKECAFQSLAHLOYHORVHTGKEYPCI ECKKAFSDGSYLVOHORLHSGKRPYECKECCAFQSLAHLOYHORVHTGKEYPC GKAFSHRSSTLHORLHTEGKEYECKECCAFQSOVAHLTLKRLHTGGERPYECKECCG AFROSVAHLAHORIHTEGESVITLSALPHYQV"
repeat_region	complement(24098..24243) /rpt_family="LlMB" repeat_region complement(24295..24317) /rpt_family="AT_rich"







\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1529: contig of 1529 bp in length  
\* 1530 1629: gap of unknown length  
\* 1630 2945: contig of 1316 bp in length  
\* 2946 3045: gap of unknown length  
\* 3046 4365: contig of 1320 bp in length  
\* 4366 4465: gap of unknown length  
\* 4466 6057: contig of 1592 bp in length  
\* 6058 6157: gap of unknown length  
\* 6158 7646: contig of 1489 bp in length  
\* 7647 7746: gap of unknown length  
\* 7747 8848: contig of 1102 bp in length  
\* 8849 9963: contig of 1015 bp in length  
\* 9964 10063: gap of unknown length  
\* 10064 11257: contig of 1194 bp in length  
\* 11258 11357: gap of unknown length  
\* 11358 12937: contig of 1580 bp in length  
\* 12938 13037: gap of unknown length  
\* 13038 14304: contig of 1267 bp in length  
\* 14305 14404: gap of unknown length  
\* 14405 15589: contig of 1185 bp in length  
\* 15590 15689: gap of unknown length  
\* 15690 16899: contig of 1210 bp in length  
\* 16900 16999: gap of unknown length  
\* 17000 18157: contig of 1158 bp in length  
\* 18158 18257: gap of unknown length  
\* 18258 19628: contig of 1371 bp in length  
\* 19629 19728: gap of unknown length  
\* 19729 21140: contig of 1412 bp in length  
\* 21141 22406: contig of 1166 bp in length  
\* 22407 22506: gap of unknown length  
\* 22507 23647: contig of 1141 bp in length  
\* 23648 23747: gap of unknown length  
\* 23748 25186: contig of 1439 bp in length  
\* 25187 25286: gap of unknown length  
\* 25287 26291: contig of 1005 bp in length  
\* 26292 26391: gap of unknown length  
\* 26392 28358: contig of 1967 bp in length  
\* 28359 28458: gap of unknown length  
\* 28460 30663: contig of 2205 bp in length  
\* 30664 30763: gap of unknown length  
\* 30764 32618: contig of 1855 bp in length  
\* 32619 32718: gap of unknown length  
\* 32719 34502: contig of 1784 bp in length  
\* 34503 34602: gap of unknown length  
\* 34603 36285: contig of 1683 bp in length  
\* 36286 36385: gap of unknown length  
\* 36386 38410: contig of 2025 bp in length  
\* 38411 38510: gap of unknown length  
\* 38511 39931: contig of 1421 bp in length  
\* 39932 40031: gap of unknown length  
\* 40032 42488: contig of 2457 bp in length  
\* 42489 42588: gap of unknown length  
\* 42589 44562: contig of 1974 bp in length  
\* 44563 44662: gap of unknown length  
\* 44663 47260: contig of 2398 bp in length  
\* 47261 47360: gap of unknown length  
\* 47361 49286: contig of 1926 bp in length  
\* 49287 49387: gap of unknown length  
\* 49388 51557: contig of 2171 bp in length  
\* 51558 51657: gap of unknown length  
\* 51659 54101: contig of 2444 bp in length  
\* 54102 54201: gap of unknown length  
\* 54202 56305: contig of 2104 bp in length  
\* 56306 56405: gap of unknown length  
\* 56406 59008: contig of 2603 bp in length

59009 59108: gap of unknown length  
\* 59109 61501: contig of 2393 bp in length  
\* 61502 61601: gap of unknown length  
\* 61602 63740: contig of 2139 bp in length  
\* 63741 63840: gap of unknown length  
\* 63841 67047: contig of 3207 bp in length  
\* 67048 67147: gap of unknown length  
\* 67148 70060: contig of 2913 bp in length  
\* 70061 70160: gap of unknown length  
\* 70161 74802: contig of 4642 bp in length  
\* 74803 74902: gap of unknown length  
\* 74903 77359: contig of 2457 bp in length  
\* 77360 77459: gap of unknown length  
\* 77460 80121: contig of 2662 bp in length  
\* 80122 80221: gap of unknown length  
\* 80222 82487: contig of 2266 bp in length  
\* 82488 82587: gap of unknown length  
\* 82588 87161: contig of 4574 bp in length  
\* 87162 87261: gap of unknown length  
\* 87262 89779: contig of 2518 bp in length  
\* 89780 89879: gap of unknown length  
\* 89880 93959: contig of 4080 bp in length  
\* 93960 94059: gap of unknown length  
\* 94060 99749: contig of 5690 bp in length  
\* 99750 99849: gap of unknown length  
\* 99850 105428: contig of 5579 bp in length  
\* 105429 105528: gap of unknown length  
\* 105529 108647: contig of 3119 bp in length  
\* 108648 108748: gap of unknown length  
\* 111451: contig of 2704 bp in length  
\* 111551: gap of unknown length  
\* 111552 117443: contig of 5892 bp in length  
\* 117444 117543: gap of unknown length  
\* 117544 121850: contig of 4307 bp in length  
\* 121851 121950: gap of unknown length  
\* 121951 128031: contig of 6081 bp in length  
\* 128032 128131: gap of unknown length  
\* 128132 132701: contig of 4570 bp in length  
\* 132702 132801: gap of unknown length  
\* 132802 137953: contig of 5152 bp in length  
\* 137954 138053: gap of unknown length  
\* 138054 146729: contig of 8676 bp in length  
\* 146730 146829: gap of unknown length  
\* 146830 152296: contig of 5467 bp in length  
\* 152297 152396: gap of unknown length  
\* 152397 157499: contig of 5103 bp in length  
\* 157500 157599: gap of unknown length  
\* 157600 163449: contig of 5850 bp in length  
\* 163450 163549: gap of unknown length  
\* 163550 168124: contig of 4575 bp in length  
\* 168125 168224: gap of unknown length  
\* 168225 178118: contig of 9894 bp in length  
\* 178119 178218: gap of unknown length  
\* 178219 189320: contig of 11102 bp in length  
\* 189321 189420: gap of unknown length  
\* 189421 196694: contig of 7274 bp in length  
\* 196695 196794: gap of unknown length  
\* 196795 206464: contig of 9670 bp in length  
\* 206465 206564: gap of unknown length  
\* 206565 217803: contig of 11239 bp in length  
\* 217804 217903: gap of unknown length  
\* 217904 230534: contig of 12631 bp in length.  
\* Location/Qualifiers  
\* 1. 230534

## FEATURES

source

/organism="Mus musculus"  
/db\_xref="taxon:10090"

/clone\_lib="RP23-86E5"

## BASE COUNT

61646 a 50860 c 50276 g 61304 t 6448 others

## ORIGIN

Query Match

9.6%; Score 283.6; DB 65; Length 230534;

Best Local Similarity 57.8%; Pred. No. 4.2e-35;  
Matches 580; Conservative 0; Mismatches 394; Indels 30; Gaps 3;

Search completed: February 22, 2001, 03:04:53  
Job time: 14351 sec

```
OY 821 cccgtgacccgcccccttccagttgtctgttggcaagccttcggcaagcccaact 880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221303 CAGGCGAGAGACCGTTCCGTGTGGAGTGGCCGCGACGTTCCAGCCGACCTCTCAC 221362

OY 881 tgatcgtccaccgcccgtgcaacagggagcccccaagtgccccagtgccgga 940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221363 TACTCGGCGACACAGCTACGATGAGGAGAGGCGCTTCTGTGGCCGACTGTGGCC 221422

OY 941 agcgcttaaccaataagccctatctgaactgcaccgagcgtccacacggcgagaagc 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221423 AGGCGTTGTGGCGAGCGCCCGCTGGAGAGACCGGCGGTGTACACAGCGGCGAGCAC 221482

OY 1001 cctaccggtgcaagagtgccgcccgttcggcaacaaccacgtctgtctaca 1060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221483 CGTTCCGCTGCGACAGAGTGTGGAGCTTTCGGGAGCGCTCCAACTTACTGACAGCAC 221542

OY 1061 gcaagattcacagagcatccgaggggttcgcccaggccgcccccgccggagagcccc 1120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221543 AGCGCATCCAGCGGATCCGCGCTGGTCCAGCGGCATTCCTGCGCCCTCGCGGAGTCC 221602

OY 1121 agctgcacgcccgcgcccaagagtlccgcccagcccaaccggcggtaccttgaac 1180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221603 CGGACCCCGCGGCCCC-----TTCCCATGACAGGAGTCTGCGAGAG 221645

OY 1181 cggcccaggagccgcccagggggccggccggccagagcaaccggagagcccgcc 1240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221646 CTTCCCGACGACGCGC-----GTGCTGTCTAGAGCACCAAGCGGTGCACACCGGAGACA 221698

OY 1241 cccctccctctacagctgcaagctgcaagagcttcggcgtggaagcgttccctgc 1300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221699 AGTC-----CTTGGAGATGCGTGGAGTGGCGAGCGCTTGGCCGCGCTTGGTGTCT 221752

OY 1301 gggcccaccagcgcaacacgagcgagcccttaacctgcgcgaagtgcgggaaga 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221753 TGCACACCGCGGCGCTGACACAGCGGCGAACGCGCTTGCCTGTGCGCGAGTCCGGCCAAA 221812

OY 1361 acttcggcaagaagagcgcacctggtgcagctccgggtgcatccggcgagagcgccct 1420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221813 GCTTCCGGCAGCGCTTAATCTCACCCAGCACACGCGCATCCACACTGGAGACCGGCTCT 221872

OY 1421 tgcgttcgagagagtgccgcccgttcccaaggagccatctggtgcgcacatcgcgc 1480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221873 TCGCCTGGCGCGAATGTGGCAAGACCTTCCGCGAGGCGCCACGCTCACGAGCACTGC 221932

OY 1481 cggaccagcccccgatcggcccttcgtgtgttcggaactgcggaagccttcgcgcaca 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221933 GTGTGCATACGGGTGAGAGGCCCTTCCGCTGCGCCGAGTGGCGCCAGCGCTTACGCCA 221992

OY 1541 aaccttaactggcgagcgacggcgcatccacacggcggaagaagcctacgtctgcgcg 1600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221993 GACTTAAGCTACGCGGCCACCAAGACCCATATCTGGCGAGAGCCCTACTGTGGGTG 222052

OY 1601 actgcgcaaaagccttcagcagaagttccaactgtgtgcagcagcgcgcatccacaag 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222053 AGTGGGCGCTCGGCTTCAAGCGAGGTGTCTCGGCTCACGAGACCAAGCGCATCCACACAG 222112

OY 1661 gcgagcgcccttaagcgttccggaactgcgaactgcgaacttcagcagaagttccaactca 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222113 GTGACCGGCGCTTCCGCTGTCCGATGTGGCAGAGTTTCCGCGAGCGACGCCAACCTCA 222172

OY 1721 tcaaccagcgcaagagcacaatccgggagcgcccttcgtgtgtccatctgtgcccaga 1780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222173 CCGACACCGCGCGCATCCATTAAGGCGAGCGACCTTACGCGGTCTCCGATGTGGCAAG 222232

OY 1781 ccttcgacgacgagagagactcctggtccacacagaagaagcac 1824
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222233 CCTTTCGCGAGGCGCCACGCTACAGCAGCACTGCGCACTCAC 222276
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2001, 03:04:53 ; Search time 7860.64 Seconds  
(without alignments)  
246.101 Million cell updates/sec

Title: US-09-477-392-3

Perfect score: 378  
Sequence: 1 ggtgagatgcgcgcgcacg.....tacctctgaaccgcccacg 378

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBml:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pr1:\*  
9: gb\_pr2:\*  
10: gb\_pr3:\*  
11: gb\_ro:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: em\_fun:\*  
15: em\_hum1:\*  
16: em\_hum2:\*  
17: em\_in:\*  
18: em\_om:\*  
19: em\_or:\*  
20: em\_ov:\*  
21: em\_pat:\*  
22: em\_ph:\*  
23: em\_pl:\*  
24: em\_ro:\*  
25: em\_sy:\*  
26: em\_sy:\*  
27: em\_un:\*  
28: em\_vl:\*  
29: gb\_htg1:\*  
30: gb\_htg2:\*  
31: gb\_in1:\*  
32: gb\_in2:\*  
33: em\_ba1:\*  
34: em\_ba2:\*  
35: em\_hum3:\*  
36: em\_hum4:\*  
37: gb\_pr4:\*  
38: gb\_htg3:\*  
39: gb\_htg4:\*  
40: gb\_htg5:\*  
41: gb\_htg6:\*  
42: gb\_htg7:\*  
43: em\_htg1:\*

44: em\_htg2:\*  
45: em\_htg3:\*  
46: em\_hum5:\*  
47: gb\_pl3:\*  
48: gb\_pr5:\*  
49: gb\_htg8:\*  
50: gb\_htg9:\*  
51: gb\_htg10:\*  
52: gb\_htg11:\*  
53: gb\_htg12:\*  
54: gb\_htg13:\*  
55: gb\_in3:\*  
56: gb\_in3:\*  
57: gb\_htg15:\*  
58: gb\_htg16:\*  
59: gb\_htg17:\*  
60: em\_htg4:\*  
61: em\_htg5:\*  
62: em\_htg6:\*  
63: em\_htg7:\*  
64: em\_hum6:\*  
65: gb\_htg18:\*  
66: gb\_htg19:\*  
67: gb\_htg20:\*  
68: gb\_htg21:\*  
69: gb\_htg22:\*  
70: gb\_htg23:\*  
71: gb\_vl1:\*  
72: gb\_vl2:\*  
73: gb\_ba3:\*  
74: em\_htg8:\*  
75: em\_htg9:\*  
76: em\_htg10:\*  
77: gb\_pr6:\*  
78: gb\_pr7:\*  
79: gb\_sts1:\*  
80: gb\_sts2:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_htg0:\*  
84: gb\_htg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378	100.0	2878	77 AK025356	AK025356 Homo sapi
2	378	100.0	2878	77 HSA245553	AJ245553 Homo sapi
3	378	100.0	2954	37 AF201303	AF201303 Homo sapi
4	378	100.0	132150	9 AC005586	AC005586 Homo sapi
5	243	64.3	201458	40 AC015887	AC015887 Mus muscu
6	142.6	37.7	759	78 HSM802596	AL162065 Homo sapi
7	126.6	33.5	210923	39 AC013273	AC004877 Homo sapi
8	126.6	33.3	2094	11 MM2P29	X55126 M. musculus
9	126	33.2	162570	58 AC073314	AC073314 Homo sapi
10	125.4	32.3	5627	8 AB011129	AB011129 Homo sapi
11	122.2	32.3	154218	10 AC020663	AC020663 Homo sapi
12	122.2	31.8	138278	84 HSI599615	AL109966 Homo sapi
13	120.2	31.4	29802	29 AC007518	AC007518 Mus muscu
14	118.8	31.0	201458	8 AC015887	AC015887 Mus muscu
15	117	30.9	6045	8 AB002324	AB002324 Human mRN
16	116.8	30.9	185306	39 AC013570	AC013570 Homo sapi
17	116.8	30.7	130067	9 AC007228	AC007228 Homo sapi
18	116.6	30.7	1554	37 AF269249	AF269249 Homo sapi
19	116.2	30.5	2330	37 AK023989	AK023989 Homo sapi
20	115.2	30.5	3000	10 AF060503	AF060503 Homo sapi
21	115.2	30.5	3000	10 AF060503	AF060503 Homo sapi

22	115	30.4	1909	37	AF242768	AF242768 Homo sapi
23	114.4	30.3	1824	10	AF000560	AF000560 Homo sapi
24	114.4	30.3	164884	51	AC023887	AC023887 Homo sapi
25	114.4	30.3	166803	55	AC048382	AC048382 Homo sapi
26	114.4	30.3	194486	41	AC016771	AC016771 Homo sapi
27	114.4	30.3	200676	51	AC024580	AC024580 Homo sapi
28	113.8	30.1	2454	11	AB041616	AB041616 Homo sapi
29	113.8	30.1	259451	40	AC015605	AC015605 Mus muscu
30	113.6	30.1	183525	39	AC012313	AC012313 Homo sapi
31	111.4	29.5	2692	37	AK024191	AK024191 Homo sapi
32	111.4	29.5	124237	66	AF188001	AF188001 Homo sapi
33	110.4	29.2	124237	66	AF189001	AF189001 Homo sapi
34	109.4	28.9	259451	40	AC015605	AC015605 Mus muscu
35	108.8	28.8	2587	10	AF055077	AF055077 Homo sapi
36	108.8	28.8	2678	85	H00M2E1	M58297 Human zinc
37	108.8	28.8	2905	10	AF055078	AF055078 Homo sapi
38	108.8	28.8	12456	37	AF161886	AF161886 Homo sapi
39	108.8	28.8	40616	41	AC016630	AC016630 Homo sapi
40	108.8	28.8	44210	41	AC016629	AC016629 Homo sapi
41	108.4	28.7	4582	11	M0SP1M2X	M98502 Mus muscu
42	108.4	28.7	203100	55	AC060772	AC060772 Mus muscu
43	108.2	28.6	113247	65	AC079506	AC079506 Mus muscu
44	108.2	28.6	168700	30	AC009115	AC009115 Homo sapi
45	108.2	28.6	230534	65	AC079575	AC079575 Mus muscu

## ALIGNMENTS

RESULT	1
AK025356	
LOCUS	AK025356 2878 bp mRNA, PRI 29-SEP-2000
DEFINITION	Homo sapiens CDNA: FLJ21703 f1s, clone COL09895, highly similar to
ACCESSION	HS4245553 Homo sapiens mRNA for AP4 zinc finger protein.
VERSION	AK025356
KEYWORDS	AK025356.1 GI:10437856
SOURCE	oligo capping; f1s (full insert sequence).
ORGANISM	Homo sapiens colon cDNA to mRNA, clone_t1b:COL clone:COL09895.
	Homo sapiens

REFERENCE	TITLE
AUTHORS	JOURNAL
Mammalia, Eutheria, Primates; Catarrhini; Hominoidea; Homo. Sensu Jolly, Petzold, Chazotac, Gannat, Verbecchia, Eucetis Comu.	
1 (sites)	
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.	
NEDO human cDNA sequencing project	
Unpublished (2000)	
2 (bases 1 to 2876)	
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	
Direct Submission	
Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumitted Sugano, Institute of Medical science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)	
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	

```

FEATURES
source
    location/Qualifiers
    1..2878
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="COL09895"
    /clone_1lb="COL"
    /tissue_type="colon"
    /note="Cloning vector pME18SFL3"
    1..2878
misc_feature

```

[illegible]

```

RESULT 2
HSA245553
LOCUS HSA245553 2878 bp mRNA PRI 18-AUG-1999
DEFINITION Homo sapiens mRNA for AP4 zinc finger protein.
ACCESSION AJ245553
VERSION AJ245553.1 GI:5748564
KEYWORDS AP4 gene; AP4 protein; zinc finger protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2878)
AUTHORS Dobner,T.G., Fischer,M. and Groitl,P.
JOURNAL Cloning of a novel zinc finger protein
REFERENCE 2 (bases 1 to 2878)
AUTHORS Dobner,T.G.
JOURNAL Direct Submmission
SUBMITTED (17-AUG-1999) Dobner T.G., Institut fuer Medizinische
Mikrobiologie und Hygiene, Universitaet Regensburg,
Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY
FEATURES
SOURCE
1. 2878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
106..1776
/gene="AP4"
106..1776
/gene="AP4"
/feature="AP4"
/note="zinc finger protein"
/codon_start=1
/product="AP4 protein"
/protein_id="CAB53100.1"

```

```
gene
/gene="Ap4"
106..1776
CDS
106..1776
/gene="Ap4"
/note="zinc finger protein"
/codon_start=1
/product="Ap4 protein"
/protein_id="CA853100.1"
```





MEDLINE  
 99063792  
 2 (bases 1 to 132150)  
 Stoneking,T., Ozersky,P., Woldmann,P. and Le,T.  
 TITLE  
 The sequence of Homo sapiens PAC clone RP4-584D14  
 JOURNAL  
 Unpublished  
 3 (bases 1 to 132150)  
 Waterston,R.H.  
 REFERENCE  
 Direct Submission  
 Submitted (01-SEP-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108 USA  
 4 (bases 1 to 132150)  
 Waterston,R.H.  
 REFERENCE  
 Direct Submission  
 Submitted (12-JUN-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 132150)  
 Waterston,R.  
 REFERENCE  
 Direct Submission  
 Submitted (30-SEP-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jun 12, 2000 this sequence version replaced gi:3907529.  
 COMMENT  
 -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: saplens@watscn.wustl.edu  
 -----  
 Summary Statistics  
 -----  
 Center project name: H\_DJ0584D14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

#### VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584D14; actual end is at base position 131954 of RP4-584D14.

The sequence RP4-584D14 from base position 21655 to 22775 is GT rich. This region was sized with PCR from genomic DNA and the hindIII digest with band size 3643 real, and 3642 insilico. The

FEATURES	sequence RP4-584D14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.
source	1. 132150 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q31-q35" /clone="RP4-584D14" /clone_lib="RPCI-4"
repeat_region	230..315 /rpt_family="MIR"
repeat_region	960..1288 /rpt_family="MER4-group"
repeat_region	1360..1468 /rpt_family="L2"
repeat_region	1562..1616 /rpt_family="L2"
misc_feature	1733..2080 /note="similar to EST AW510685 (NID:g7148763) hc89e08.x1"
repeat_region	2051..2114 /rpt_family="L2"
misc_feature	2148..3224 /note="CpG island (%GC=70.7, o/e=0.90, #CpGs=107)"
repeat_region	2347..2426 /rpt_family="GC-rich"
repeat_region	3141..3232 /rpt_family="Alu"
repeat_region	3235..3368 /rpt_family="Alu"
repeat_region	3795..3888 /rpt_family="MIR"
repeat_region	4037..4199 /rpt_family="Alu"
repeat_region	4308..4593 /rpt_family="MER4-group"
misc_feature	4872..5309 /note="similar to EST AI827463 (NID:g5448134) wf29f03.x1"
repeat_region	4882..4909 /rpt_family="AT-rich"
repeat_region	5679..5773 /rpt_family="MIR"
misc_feature	5947..5952 /note="match to EST AA482420 (NID:g2210098) zt34f02.r1"
repeat_region	6148..6450 /rpt_family="Alu"
repeat_region	6453..6690 /rpt_family="MER4-group"
repeat_region	6695..7007 /rpt_family="Alu"
repeat_region	7008..7117 /rpt_family="MER4-group"
repeat_region	7180..7355 /rpt_family="MER4-group"
misc_feature	7615 /note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
misc_feature	7883..8397 /note="match to EST W72943 (NID:g1383235) zd54f12.r1"
misc_feature	7886..8357 /note="match to EST A1144211 (NID:g3666020) qb93g11.x1"
misc_feature	7886..8562 /note="similar to EST AW411281 (NID:g6936822) fh11d05.y1"
misc_feature	7886..8403 /note="similar to EST AW246540 (NID:g6589533)"
misc_feature	7886..8451 /note="match to EST AA129215 (NID:g1689084) zn84b02.s1"
STS	7886..8133 /db_xref="GI:1131717"
misc_feature	7886..8146 /note="similar to EST AI868891 (NID:g5542895) wc49f12.x1"
misc_feature	7888..8340 /note="match to EST H37790 (NID:g907289) yp46h05.s1"
misc_feature	7896..8372



```
* 4059 4158: gap of 100 bp
* 4159 5831: contig of 1673 bp in length
* 5832 5931: gap of 100 bp
* 5932 7380: contig of 1449 bp in length
* 7381 7480: gap of 100 bp
* 7481 8820: contig of 1340 bp in length
* 8821 8920: gap of 100 bp
* 8921 10519: contig of 1599 bp in length
* 10520 10619: gap of 100 bp
* 10620 13322: contig of 2703 bp in length
* 13323 13422: gap of 100 bp
* 13423 15194: contig of 1772 bp in length
* 15195 15294: gap of 100 bp
* 15295 18405: contig of 3111 bp in length
* 18406 18505: gap of 100 bp
* 18506 20293: contig of 1788 bp in length
* 20294 20393: gap of 100 bp
* 20394 24037: contig of 3644 bp in length
* 24038 24137: gap of 100 bp
* 24138 30849: contig of 6712 bp in length
* 30850 30949: gap of 100 bp
* 30950 44871: contig of 13922 bp in length
* 44872 44971: gap of 100 bp
* 44972 61254: contig of 16283 bp in length
* 61255 61354: gap of 100 bp
* 61355 74139: contig of 12785 bp in length
* 74140 74239: gap of 100 bp
* 74240 99150: contig of 24911 bp in length
* 99151 99250: gap of 100 bp
* 99251 129995: contig of 30745 bp in length
* 129996 130095: gap of 100 bp
* 130096 163192: contig of 33097 bp in length
* 163193 163292: gap of 100 bp
* 163293 201458: contig of 38166 bp in length.
*
* Location/Qualifiers
* 1..201458
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone_lib="CITB Mouse BAC"
  1..1060
    misc_feature /note="assembly-fragment"
    1161..2600
      misc_feature /note="assembly-fragment"
    2701..4058
      misc_feature /note="assembly-fragment"
    4159..5831
      misc_feature /note="assembly-fragment"
    5932..7380
      misc_feature /note="assembly-fragment"
    7481..8820
      misc_feature /note="assembly-fragment"
    8921..10519
      misc_feature /note="assembly-fragment"
    10620..13322
      misc_feature /note="assembly-fragment"
    13423..15194
      misc_feature /note="assembly-fragment"
    15295..18405
      misc_feature /note="assembly-fragment"
    18506..20293
      misc_feature /note="assembly-fragment"
    clone_end:r7
    vector_side:left"
    20394..24037
      misc_feature /note="assembly-fragment"
    24138..30849
      misc_feature /note="assembly-fragment"
    30950..44871
      misc_feature /note="assembly-fragment"
    44972..61254
      misc_feature /note="assembly-fragment"
    61355..74139
```

```
Query Match 64.3%; Score 243; DB 40; Length 201458;
Best Local Similarity 83.4%; Pred. No. 1.1e-34;
Matches 276; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
misc_feature /note="assembly-fragment"
74240..99150
misc_feature /note="assembly-fragment"
99251..129995
misc_feature /note="assembly-fragment"
130096..163192
misc_feature /note="assembly-fragment"
163293..201458
BASE COUNT 54496 a 46124 c 45270 g 53666 t 1902 others
ORIGIN
```

```
QY 1 ggttgagatgcgtgcagcccttcacagtgctgtgtggaagcgttcgcaac 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132529 GGTGGAGACGCTGTGACCGCCCTTCCAGTGTGCTCGCGCAAGCGTTCCGCCAC 132588
QY 61 aagcccaacttatgcctcacccgcgtgcacagggagcggcccccacatggcccc 120
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132589 AAGCCCAATCTGATGCGCCACCGCGGTGCACACTGTGTGACGACACACCAAGTCCCA 132648
QY 121 ggttcggaagcgtcttacaataagccctactctgacttcacgcgcgtccacacc 180
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132649 GAGTGGGGGAAACGTTTACCAACAAGCCCTACCTGACCTGCACCGGCGCATACATACC 132708
QY 181 ggcgaagaagccttaccgcgtgcaagaagtgcgccgcgttcgcgcacaacccaactg 240
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132709 GCGGGAAGACCTTACCAAGCACCAGTGTGCGCCCGCTTCCGCCCAAAACCAACTG 132768
QY 241 cgtctcacagaagaattcacagaagatccgagggtgsgccacgcgcgcgcgcgcgcg 300
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132769 TTGTGCAACGAAATTCACCAAGCGCTTGAGAGTCTGACGAGGTGCCCCACACCCC 132828
QY 301 gggagcccccagctgcacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 331
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132829 GAGAGTCAACCATTTGCACAGCAGAGCCTATGCG 132859
```

```
RESULT 6
HSM802596/c 759 bp mRNA PRI 23-MAR-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp762K135 (from clone DKFZp762K135).
DEFINITION A1162065
ACCESSION A1162065
VERSION A1162065.1 GI:7328138
KEYWORDS
SOURCE
ORGANISM human.
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 759)  
Bloecher, H., Boeher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.  
Direct Submission  
Submitted (15-MAR-2000) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp762K135) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

```
FEATURES
SOURCE 1..759
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/clone="DKFZp762K135"
/clone_id="762 (synonym: hme12). Vector psport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/issue_type="melanoma (Meko cell line)"
BASE COUNT      117 a      246 c      252 g      144 t
ORIGIN

```

```

Query Match      37.7%; Score 142.6; DB 78; Length 759;
Best Local Similarity 67.9%; Pred. No. 4.3e-16;
Matches 199; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 8 atgcgctgacggccgcccctgctgtgtgtgcaagcgttcgagcacaagccca 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 709 ACCGCCCGGATGCGCCCTTCTGTGTCCGACTGCGGCAAGGCTTCCGCACAAACCTT 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 actgctgctcaccgctgctgcaacgagcgagcgagccaccagctgcccagtgctg 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 ACCGCGGGGGGCGGCGGCGCATCCACACCGGCGAAGACCTTACGCTGTGCCCGACTCG 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 ggaagcgttaacataagccctatctgacttcgacgagcgatccacacgagcgaga 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 GCAAGGCTTCAGCGCAGAAATCCAAACCTGTGTGCGACCGCGCATCCACACGCGCGAGC 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 188 agccctaccgctgcaagagtgctgagcgagcgagcgagcacaacaccactgtctgtc 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 529 GGGCTTACGCTGTCCGACGCGACCGCGCTTCAGCGCAAGATCCAACTCATCACC 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 248 acagcaagatcacaagcgatccgagcgagcgagcgagcgagcgagcgagcgagcg 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 ACCGCAAGAGGCATCCGCGGAGCGCGCTTCTGTGTGCGCATCTGTGTGCGCAG 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
AC004877/c      128361 bp      DNA      PRI      21-DEC-1999
LOCUS      Homo sapiens PAC clone RP4-751H13 from 7q35-qter, complete
DEFINITION
sequence.
ACCESSION      AC004877
VERSION        AC004877.1 GI:3638954
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 128361)
AUTHORS      Leonard, S., Graves, T. and Strommatt, C.
TITLE        The sequence of Homo sapiens PAC clone RP4-751H13
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 128361)
AUTHORS      Waterston, R. H.
TITLE        Direct Submission
JOURNAL      Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      3 (bases 1 to 128361)
AUTHORS      Waterston, R.
TITLE        Direct Submission
JOURNAL      Submitted (19-SEP-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      4 (bases 1 to 128361)
AUTHORS      Waterston, R.
TITLE        Direct Submission
JOURNAL      Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 19, 1998 this sequence version replaced gi:3213120.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapliens@wustl.wustl.edu
----- Summary Statistics

```

Center project name: H\_DV0751H13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTR/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP4-81M16, 200 bp overlap. Actual start of this clone is at base position 1 of RP4-751H13; actual end is at 128361 of RP4-751H13.

#### FEATURES

##### source

```

1. 128361
   location/Qualifiers
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7q35-qter"
   /clone="RP4-751H13"
   /clone_id="unknown"
   20..155
   /rpt_family="Alu"
   repeat_region
   860..941
   /rpt_family="MER1-type?"
   2531..2885
   /rpt_family="Retroviral"
   2924..3246
   /rpt_family="Alu"
   repeat_region
   3300..3727
   /rpt_family="Retroviral"
   4146..4287
   /rpt_family="L2"
   5253..5728
   /rpt_family="MALR"
   5933..6243
   /rpt_family="Alu"
   repeat_region
   7787..7870
   /rpt_family="MIR"
   7798..8145
   /note="match to EST AA084720 (NID:g1626811) zn06f02.s1"
   repeat_region
   8259..8311
   /rpt_family="L2"
   9100..9426
   /note="match to EST AA080867 (NID:g1623560) zn06f02.r1"
   repeat_region
   9773..10086
   /rpt_family="Alu"
   complement(9947..10370)
   /note="match to EST AA828832 (NID:g2901931) od80c03.s1"
   9947..10370
   /note="match to EST AA828832 (NID:g2901931) od80c03.s1"
   10066..10388
   /note="match to EST AA857387 (NID:g2945689) od83h10.s1"
   repeat_region
   10287..10734
   /rpt_family="MALR"
   11605..12220
   /rpt_family="L1"
   12222..12615
   repeat_region

```

repeat_region	/rptL_family="Retroviral"
repeat_region	12617..12755
repeat_region	/rptL_family="L1"
repeat_region	12768..12890
repeat_region	/rptL_family="MIR"
repeat_region	12891..13196
repeat_region	/rptL_family="Alu"
repeat_region	13197..13209
repeat_region	/rptL_family="MIR"
repeat_region	13314..13618
repeat_region	/rptL_family="Alu"
repeat_region	13693..14004
repeat_region	/rptL_family="Alu"
repeat_region	14040..14162
repeat_region	/rptL_family="MIR"
repeat_region	15625..15931
repeat_region	/rptL_family="Alu"
repeat_region	16550..16702
repeat_region	/rptL_family="L1"
repeat_region	18090..19036
repeat_region	/rptL_family="L12"
repeat_region	19479..19798
repeat_region	/note="match to EST AA939299 (NID:g3099212) ol78g08.s1"
repeat_region	19479..19656
repeat_region	/note="similar to EST H89462 (NID:g1079892)"
repeat_region	19817..20311
repeat_region	/note="match to EST AA625192 (NID:g2537577) af67h06.r1"
repeat_region	19860..21428
repeat_region	/note="CPG island (%GC=70.4, o/e=0.70, #CGs=171)"
repeat_region	complement(19861..20173)
repeat_region	/note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
repeat_region	19861..20239
repeat_region	/note="similar to EST AA923119 (NID:g3070428) ok91e12.s1"
repeat_region	complement(19898..20090)
repeat_region	/note="match to EST AA771702 (NID:g2823485) ai32a10.s1"
repeat_region	complement(19937..20235)
repeat_region	/note="match to EST AA885789 (NID:g3000897) oj35a06.s1"
repeat_region	complement(20009..21523)
repeat_region	/gene="W0GSC:H.DJ0751H13.3"
repeat_region	complement(20009..21523)
repeat_region	/gene="W0GSC:H.DJ0751H13.3"
repeat_region	/note="zinc finger-like; similar to P52742 (PID:g1731411);
repeat_region	H.DJ0751H13.3"
repeat_region	/codon_start=1
repeat_region	/evidence=not_experimental
repeat_region	/protein_id="AAC36300.1"
repeat_region	/db_xref="GI:3638956"
repeat_region	/translation="MIRKYKEDDEDEQEAEEVEWPHQLSLPSPAPDLGHLAAVY
repeat_region	LEPGAGALSGALSGMGPMPKPYGCGCEGERFSDOLTLRLHRIHNGEERPYCAECG
repeat_region	RSQFLGQAHMLHQRSHGGERPFCSECDKRRFQKAKHLTRLRLHNGEERPYCAECG
repeat_region	PSGFLGQAHMLHQRSHGGERPFCSECDKRRFQKAKHLTRLRLHNGEERPYCAECG
repeat_region	THKQLVKQRVHQTAGPARSPDSSAPSHSTASPTSPFSGPKRFACSDGLSGFWK
repeat_region	KNLALHQCIRHESGRPFSCDECALGATVDAAPAKPLAAGPGGCGSDVVAQRAAE
repeat_region	SGSERFPCDPCGRFSRSHLGRHVAATGTSRPHACVACARSFSKTNLVHQAITHGS
repeat_region	AARFACACGGRFSRSHLGRHVAATGTSRPHACVACARSFSKTNLVHQAITHGS
repeat_region	PFSPCGCGSFRSKTNLVHQAITHGEAAHAAPADALAAPAMSAPPEVAPPLPF"
repeat_region	complement(20014..20272)
repeat_region	/gene="W0GSC:H.DJ0751H13.3"
repeat_region	/note="match to EST AA771702 (NID:g2823485) ai32a10.s1"
repeat_region	complement(20059..20239)
repeat_region	/gene="W0GSC:H.DJ0751H13.3"
repeat_region	/note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
repeat_region	complement(20097..20319)
repeat_region	/gene="W0GSC:H.DJ0751H13.3"
repeat_region	/note="match to EST AA885789 (NID:g3000897) oj35a06.s1"
repeat_region	20176..20322
repeat_region	/note="match to EST AA894411 (NID:g3030812) oe79b04.s1"
repeat_region	20298..20740
repeat_region	/note="similar to Mus musculus EST AAO21873 (NID:g1485629)
repeat_region	mb85f12.r1
repeat_region	20534..20724
repeat_region	/note="match to EST AA044857 (NID:g1523196) zf54d12.r1"
repeat_region	20819..21219

```

repeat_region      /note="match to EST: AA757828 (NID:g2805691) zg44d02.s1"
                    22551. .22486
repeat_region      /rpt_family="MIR"
                    22848. .23146
repeat_region      /rpt_family="Alu"
                    23679. .23794
repeat_region      /rpt_family="MIR"
                    23773. .23846
repeat_region      /rpt_family="L2"
                    24894. .24982
repeat_region      /rpt_family="L2"
                    24983. .25271
repeat_region      /rpt_family="Alu"
                    25272. .25616
repeat_region      /rpt_family="L2"
                    26933. .29260
misc_feature        /note="CpG_Island (%GC=73.8, o/e=0.90, #CpGs=195)"

```

Query Match	33.5%;	Score 126.6;	DB 8;	Length 128361;
Best Local Similarity	67.8%;	Pred. No. 3.7e-14;		
Matches 177; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0;

QY	11	ccgtgagacgccccttcagttgtcgtgttgggaagcgttccggacaagcccaact	70
Db	21081	CGGGGAGACGGGCCCTACCCGTGCGGGGAGTGGGGCAAGCCTTACGCCAAGAAAGATCCACC	21022
QY	71	tgatcgctcaaccgcgcgtgtgcaacagcgagcgagccccaacagtcgccgagtcgggga	130
Db	21021	TGGGGTCCGACCAAAAGACCCACCGCGGAGACGGGCCCTTCCCTGACAGGAATGGAGA	20962
QY	131	agcgctttaccatataagaacctctgactctgacacgcggagatccacacggcgagaagc	190
Db	20961	AGCGCTTTCCGCAAGACGCACTTGATTTCGGACACAGGCGCATTCATACGGGGGAGAGCG	20902
QY	191	ccctaccgctlgcaaaagatgctgcgcgcgtcttcggcgacacaaccccaactgtctgttcaca	250
Db	20901	CCCTACCACTGGCGCAACAGTGGCGACGCACACTTCACGCACAAGCAGCACTTGATGGCGCACC	20842
QY	251	gcaagattcacagatgcg	271
Db	20841	AAAGGTGCACCAAGACGGCGC	20821

RESULT	8
AC013273/c	
LOCUS	AC013273 210923 bp DNA HTG 15-JUL-2000
DEFINITION	Homo sapiens chromosome 7 clone RP11-428D5, WORKING DRAFT SEQUENCE,
ACCESSION	AC013273
VERSION	AC013273.5 GI:9211421
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 210923)
JOURNAL	Waterston,R.H.
REFERENCE	The sequence of Homo sapiens clone unpublished
AUTHORS	2 (bases 1 to 210923)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
COMMENT	Submitted (05-NOV-1999) Genome Sequencing Center, Washington MO 63108, USA On Jul 15, 2000 this sequence version replaced gi:8990985.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 Project Information  
 Center project name: H\_NH0428005

## Sequencing vector: M13; 748

NOTE: This is a working draft sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

**Source**

	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="RP11-428D5"
	1..2429
misc_feature	/note="assembly_name:Contig8"
	2530..4617
misc_feature	/note="assembly_name:Contig9"
	4718..11807
misc_feature	/note="assembly_name:Contig10"
	11908..20464
misc_feature	/note="assembly_name:Contig11"
	20565..29298
misc_feature	/note="assembly_name:Contig12"
	clone_end:"T7"
	vector_side:"right"
	29399..45700
misc_feature	/note="assembly_name:Contig13"
	43801..63662
misc_feature	/note="assembly_name:Contig14"
	clone_end:"SP6"
	vector_side:"left"
	63763..104507
misc_feature	/note="assembly_name:Contig15"
	104608..150643
misc_feature	/note="assembly_name:Contig16"
	150744..210923
misc_feature	/note="assembly_name:Contig17"
	/note="assembly_name:Contig17"
BASE COUNT	54201 a 51422 c 51731 g 52663 t
ORIGIN	906 others

Query Match	33.5%;	Score 126.6;	DB 39;	Length 210923;
Best Local Similarity	67.8%;	Pred. No. 3.1e-14;		
Matches 177; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0;

OY	11	ccgttcgaccgcgcgccttcacagctgtgcctgtttgtgcaagcgtctccgacaaagcccaact	70
Db	98278	CGGGGAGACGGGCCCTTACCCGTGCGGGAGATGGGGCAAGGCTTTCAGCCAGAAAGATCCACC	982139
OY	71	tgatcgtaaccgcgcgcgtgtgcaacagygagcagcgcaccaacagctgcccagtgcgga	130
Db	98218	TGGGCTTCCACACAAAAGACCACCGCGGAGCGGCGCTTCCCTGTGCAGAGGAATCGGAGA	98158
OY	131	agcgcttaccacataaagccctctctgactctgcacgcgagcatccacacgcgcgaagaac	190
Db	98158	AAGCGTTTTCGCGAAGAGACGCCTTGATTTGGACACCAAGCGCATTCATACGGGCGAAGGC	98099
OY	191	ccctaccgcttgcaaaagatgtgcgcgcgcgttcctgcgcacaaaccaaacctgtctgtcaaa	250
Db	98098	CCCTACCACTGGCGACAGTGCAGCAGCAGCTTCCACGCACAAAGCAGACACTTGTTGGCGCACCC	98039
OY	251	gcaagatcatcaagcagatcgcg	271
Db	98038	AAAGGCTGCACACAGCGGCCG	98018

RESULT	9
MWZFP29	
LOCUS	MWZFP29 2094 bp mRNA ROD 13-APR-1997
DEFINITION	M.musculus zfp-29 gene for zinc finger protein.
ACCESSION	X55126
VERSION	X55126.1 GI:55470
KEYWORDS	spermatogenesis; zfp-29 gene; zinc-finger.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	1 (bases 1 to 2094)
AUTHORS	Denny, P.
TITLE	Direct Submission
JOURNAL	Submitted (03-NOV-1990) Denny P., Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, LONDON, SW3 6JB, UK
REFERENCE	2 (bases 1 to 2094)
AUTHORS	Denny, P. and Ashworth, A.
TITLE	A zinc finger protein-encoding gene expressed in the post-meiotidic phase of spermatogenesis
JOURNAL	Gene 106 (2), 221-227 (1991)
MEDLINE	92039080
FEATURES	Location/Qualifiers
SOURCE	1. -2094

```

URES
source
Location/Qualifiers
1..2094
/organism="Mus musculus"
/strain="ssp. domesticus"
/db_xref="taxon:10090"
/dev_stage="adult male"
/tissue_type="testis"
/clone.lib="lambda zap cdna"
104..1948
/gene="zfp-29"
104..1948
/gene="zfp-29"
104..1948
/codon_start=1
/protein_id="CAA38920.1"
/db_xref="GI:55471"
/db_xref="SWISS-PROT:Q07230"
/translucation="MAEVPASTPLSLVLPVQPEEDEQAEVTTIMLEDAMVQEAVALI
QEDGESSEFPSSAGKSGPOEDAEAGPQALVPRRELCRMRLRPEVTKQDMITVLLE
RIQIWLMLHEHRESEEAVALVEDLTQTFRHSDPEIIOSENGENSMDMEFEVESIGMGM
LINSIGEGGOOGSDGSDPERPDGSGGAGGHAAGEDPRVPSSEGVGOLIGLOGITLGLD
EKRYPCPGCKFKFRKSHLITHERPTHGEEKYKODECGKSPDGSNFSRHOTHTHGEK
PYKCDCKCKSPRSANLITHORITHGEEKFPQACBEKGSFSPNLIARORHTHGEKPP
SCPEKCGKFSNLSINTHQGITGEEKFPACKEGGSFYSNLIKHORITHGEEKPYKCGK
TECGKGFQSSALLIHRHTHGEKPKYCEGCKGNRSRNSRLIHRHTHGEKPYKCGK

```





	RESULT	11
	ABO11129	
LOCUS	Homo sapiens mRNA for KIAA0557 protein, partial cds.	PRI 10-APR-1998
ACCESSION	ABO11129	
VERSION	ABO11129.1 GI:3043637	
KEYWORDS	KIAA0557 protein.	
SOURCE	Homo sapiens male brain cDNA to mRNA, clone_1lb:pbluescriptII SK plus clone:H11334.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
AUTHORS	Primates; Catarrhini; Homnidae; Homo.	
TITLE	1 (bases 1 to 5627)	
JOURNAL	Ohara,O., Nagase,T. and Ishikawa,K. Submitted (13-FEB-1998) to the DBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914) 2 (sites)	
REFERENCE	Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H., Miyaura,N. and Ohara,O. Prediction of the coding sequences of unidentified human genes. IX The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro DNA Res. 5 (1), 31-39 (1998)	
AUTHORS		
JOURNAL		
MEDLINE		
FEATURES	98290545	
SOURCE	Location/Qualifiers 1..5627 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="H11334" /clone_1lb="pbluescriptII SK plus" /sex="male" /tissue_type="brain" 1..1483 /gene="KIAA0557" <1..1483 /gene="KIAA0557" /codon_start=2 /product="KIAA0557 protein" /protein_id="BAA25483.1" /db_xref="GI:3043638" translation:"TIRPEKEGQPEAMATVPGIOLPTLEBDLDEELLTVKYEEPD CLEEPSEYETDPSPEPFRROLRFCTIOEVAGPPDAISRKLWELCCRWLPRLRKEO LELLLDGLFVLPLGEIQARVREDOQSPESGAVALVEGLQPKPRKHROGSSELSDDE VPIDIGGFQLKHOEAQPEDLSLEEAFSSQOPPAOLSHPOGPIELMRGPBPAP HOMASASPFLSANSOVNVEDVAAYVLSGEEPICMDPADAPLENGPGIOLEDGG DGRRDAPLRMEYRVILSARCGPGHPILVGORPAVRGIVRPDDRGPGPRRSASHG DKPYTCDECGKGFSTKSLTHNQHTHGGERXYKLCYCKSGSDSNSTHORVYTGEK PYTPCEGKRFSQSLSLVITHRTSHGSEPIACTCQGKFNNSHFSAHRKHTTGKPP TTPPCGRGFRGTDLHKHQRTHMGASLPTLQPVAPGPGAKA"	
BASE COUNT	1173 a 1606 c 1690 g 1158 t	
ORIGIN		
Query Match	32.3%; Score 122.2; DB 8; Length 5627;	
Best Local Similarity	60.1%; Pred. No. 7.6e-13;	
Matches 223; Conservative	0; Mismatches 143; Indels 5; Gaps 1;	
DB	996 ATGGGGCTGACAGCCGTTACACCTGCCCGAATTGGCAAAGCTTCAACAAGACGTCC 1055	
OY	8 atgcgctgacggccgcccttcacagtgtgacctgtgttggaaggcgttcggcacaaacca 67	
DB	996 ATGGGGCTGACAGCCGTTACACCTGCCCGAATTGGCAAAGCTTCAACAAGACGTCC 1055	
OY	68 actgatcgctacacgcgcgcgtgcacacacgcgcgcgcgcacacagtagtccccgagtgc 127	
DB	1056 ACTTACACCAAGACGACGACACACACACGAGCGGACGCGCTTACAAAGTGCCTAGTCTGG 1115	
OY	128 ggaagcgtttaccataagccctatctgacttcgcaacgcgcacatccaacacgcgcaga 187	

[illegible]

repeat\_region /rpt\_family="Alu"  
3187. .3504  
repeat\_region /rpt\_family="Alu"  
complement(3571. .3890)  
repeat\_region /rpt\_family="Alu"  
complement(3859. .4163)  
repeat\_region /rpt\_family="Alu"  
4198. .4503  
repeat\_region /rpt\_family="Alu"  
4826. .5155  
repeat\_region /rpt\_family="Alu"  
5749. .6045  
misc\_feature /rpt\_family="Alu"  
complement(6848. .6983)  
repeat\_region /note="GRAIL 2 excellent exon, frame 1"  
7473. .7776  
repeat\_region /rpt\_family="Alu"  
7931. .8385  
repeat\_region /rpt\_family="Alu"  
8808. .9139  
repeat\_region /rpt\_family="Alu"  
9394. .9682  
repeat\_region /rpt\_family="Alu"  
9515. .9812  
repeat\_region /rpt\_family="Alu"  
9702. .9987  
repeat\_region /rpt\_family="Alu"  
9965. .10332  
repeat\_region /rpt\_family="Alu"  
complement(10403. .10835)  
repeat\_region /rpt\_family="Alu"  
complement(10801. .10977)  
repeat\_region /rpt\_family="Alu"  
11580. .11780  
repeat\_region /rpt\_family="MER42"  
11803. .12101  
repeat\_region /rpt\_family="Alu"  
12282. .12576  
repeat\_region /rpt\_family="Alu"  
12643. .13073  
repeat\_region /rpt\_family="Alu"  
12923. .13237  
repeat\_region /rpt\_family="Alu"  
13420. .13725  
repeat\_region /rpt\_family="Alu"  
13761. .14148  
repeat\_region /rpt\_family="Alu"  
14072. .14371  
repeat\_region /rpt\_family="Alu"  
complement(14401. .14696)  
repeat\_region /rpt\_family="Alu"  
15061. .15366  
repeat\_region /rpt\_family="Alu"  
15383. .15706  
repeat\_region /rpt\_family="Alu"  
complement(16083. .16185)  
repeat\_region /rpt\_family="L1"  
complement(16192. .16347)  
repeat\_region /rpt\_family="Alu"  
16559. .16831  
repeat\_region /rpt\_family="Alu"  
complement(18420. .18718)  
repeat\_region /rpt\_family="Alu"  
complement(18518. .18917)  
repeat\_region /rpt\_family="Alu"  
complement(18754. .19036)  
misc\_feature /rpt\_family="Alu"  
complement(19456. .19559)  
repeat\_region /note="GRAIL 2 excellent exon, frame 1"  
20239. .20397  
repeat\_region /rpt\_family="MER20"  
complement(20808. .21104)  
repeat\_region /rpt\_family="Alu"

repeat\_region complement(22116. .22443)  
misc\_feature /rpt\_family="Alu"  
complement(22789. .22948)  
repeat\_region /note="GRAIL 2 excellent exon, frame 1"  
23346. .23646  
repeat\_region /rpt\_family="Alu"  
complement(24046. .24341)  
repeat\_region /rpt\_family="Alu"  
25205. .25461  
repeat\_region /rpt\_family="Alu"  
25605. .25862  
repeat\_region /rpt\_family="L1"  
25914. .25981  
repeat\_region /rpt\_family="Alu"  
complement(26044. .26355)  
repeat\_region /rpt\_family="Alu"  
28437. .28732  
repeat\_region /rpt\_family="Alu"  
30597. .30772  
misc\_feature /note="GRAIL 2 excellent exon, frame 2"  
31300. .31655  
repeat\_region /rpt\_family="Alu"  
32059. .32124  
repeat\_region /rpt\_family="Alu"  
complement(32302. .32492)  
repeat\_region /rpt\_family="Alu"  
32507. .32839  
repeat\_region /rpt\_family="Alu"  
33645. .34040  
repeat\_region /rpt\_family="Alu"  
complement(34202. .34278)  
repeat\_region /rpt\_family="LTR37B"  
34279. .34556  
repeat\_region /rpt\_family="Alu"  
35017. .35495  
repeat\_region /rpt\_family="Alu"  
complement(36639. .36991)  
repeat\_region /rpt\_family="Alu"  
37667. .37808  
misc\_feature /note="GRAIL 2 excellent exon, frame 0"  
complement(37920. .37961)  
repeat\_region /rpt\_family="MIR"  
38430. .38745  
repeat\_region /rpt\_family="Alu"  
38789. .38899  
repeat\_region /rpt\_family="MER3"  
complement(40939. .41089)  
repeat\_region /rpt\_family="Alu"  
42294. .42589  
repeat\_region /rpt\_family="Alu"  
complement(43093. .43396)  
repeat\_region /rpt\_family="Alu"  
complement(43274. .43695)  
repeat\_region /rpt\_family="Alu"  
44405. .44482  
repeat\_region /rpt\_family="MLTIC"  
44515. .44693  
repeat\_region /rpt\_family="Alu"  
45916. .46254  
misc\_feature /rpt\_family="Alu"  
complement(46300. .46347)  
repeat\_region /note="GRAIL 2 excellent exon, frame 1"  
46426. .46596  
repeat\_region /rpt\_family="MER20"  
47139. .47448  
repeat\_region /rpt\_family="Alu"  
47535. .47943  
repeat\_region /rpt\_family="rHE1"  
complement(47939. .48377)  
repeat\_region /rpt\_family="Alu"  
48431. .48778  
repeat\_region /rpt\_family="L1"  
48877. .49260





```

----- Project Information
Center project name: I742
Center clone name: 315_E6
----- Summary Statistics
Sequencing vector: M13; M77815; 44% of reads
Sequencing vector: Plasmid; n/a; % 0.0% of reads
55.555555555556Chemistry: Dye-terminator Big Dye; 42% of reads
Chemistry: Dye-terminator Big Dye; 42% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 192611 bases at least Q40
Consensus quality: 196150 bases at least Q30
Consensus quality: 197810 bases at least Q20
Insert size: 182000; agarose-ff
Insert size: 199558; sum-of-ctngs
Quality coverage.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1060: contig of 1060 bp in length
*
* 1061 1160: gap of 100 bp
*
* 1161 2600: contig of 1440 bp in length
*
* 2601 2700: gap of 100 bp
*
* 2701 4058: contig of 1358 bp in length
*
* 4059 4158: gap of 100 bp
*
* 4159 5831: contig of 1673 bp in length
*
* 5832 5931: gap of 100 bp
*
* 5932 7380: contig of 1449 bp in length
*
* 7381 7480: gap of 100 bp
*
* 7481 8820: contig of 1340 bp in length
*
* 8821 8920: gap of 100 bp
*
* 8921 10519: contig of 1599 bp in length
*
* 10520 10619: gap of 100 bp
*
* 10620 13322: contig of 2703 bp in length
*
* 13323 13422: gap of 100 bp
*
* 13423 15194: contig of 1772 bp in length
*
* 15195 15294: gap of 100 bp
*
* 15295 18405: contig of 3111 bp in length
*
* 18406 18505: gap of 100 bp
*
* 18506 20293: contig of 1788 bp in length
*
* 20294 20393: gap of 100 bp
*
* 20394 24037: contig of 3644 bp in length
*
* 24038 24137: gap of 100 bp
*
* 24138 30848: contig of 6712 bp in length
*
* 30850 30949: gap of 100 bp
*
* 30950 44871: contig of 13922 bp in length
*
* 44872 44971: gap of 100 bp
*
* 44972 61254: contig of 16283 bp in length
*
* 61255 61354: gap of 100 bp
*
* 61355 74133: contig of 12785 bp in length
*
* 74140 74239: gap of 100 bp
*
* 74240 99150: contig of 24911 bp in length
*
* 99151 99250: gap of 100 bp
*
* 99251 129995: contig of 30745 bp in length
*
* 129996 130095: gap of 100 bp
*
* 130096 163192: contig of 33097 bp in length
*
* 163193 163292: gap of 100 bp
*
* 163293 201458: contig of 38166 bp in length.
*
FEATURES
source
1. 201458
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="CT7-315E6"
/clone_id="CITB Mouse BAC"
1. 1060
/note="assembly_fragment"
1161. 2600
/note="assembly_fragment"
2701. 4058

```

```

misc_feature /note="assembly-fragment" 4159. 5831
misc_feature /note="assembly-fragment" 5932. 7380
misc_feature /note="assembly-fragment" 7481. 8820
misc_feature /note="assembly-fragment" 8921. 10519
misc_feature /note="assembly-fragment" 10620. 13322
misc_feature /note="assembly-fragment" 13423. 15194
misc_feature /note="assembly-fragment" 15295. 18405
misc_feature /note="assembly-fragment" 18506. 20293
misc_feature /note="assembly-fragment
clone_end:r7
vector_side:left"
20394. 24037
misc_feature /note="assembly-fragment" 24138. 30849
misc_feature /note="assembly-fragment" 30950. 44871
misc_feature /note="assembly-fragment" 44972. 61254
misc_feature /note="assembly-fragment" 61355. 74139
misc_feature /note="assembly-fragment" 74240. 99150
misc_feature /note="assembly-fragment" 99251. 129995
misc_feature /note="assembly-fragment" 130096. 163192
misc_feature /note="assembly-fragment" 163293. 201458
misc_feature /note="assembly-fragment"
BASE COUNT 54496 a 46124 c 45270 g 53666 t 1902 others
ORIGIN

```

Query Match	31.0%	Score 117	DB 40	Length 201458
Best Local Similarity	62.5%	Pred. No. 1.5e-12		
Matches 183	Conservative	0	Mismatches 110	Indels 0
				Gaps 0
QY 4	ggaagatgcgcgtgcagccgcccccttcagatgtgcctgtttgtgcgaagcgcttcgcgcaag	63		
Db 71332	ggggcctggccgcgcgacgccccttcgtgtgttcccgagttggcgcaagaccttccagctcaaa	71273		
QY 64	cccaacttgatcgtctcaaccgcgcgtgtgacacagggcgagcgagccccacagtgccccgag	123		
Db 71272	cacacacgtgcgaaggtgtcacacaggcagcagcacacggcgagcgcccttcgccctcccgag	71213		
QY 124	tgcgggaagcgcttaccacataagccctatctgacttcgcacgcgycgcataccacccgagc	183		
Db 71212	tgccgcccctgcttccagctcacaagaaatctgtctacacgacacacagcgacatccacagcgcc	71153		
QY 184	gagaagccctaccgcgtgtgcaaaagatgtgcgcgcgttcctgcgcacaaaccaactgtctg	243		
Db 71152	gagaagcccgccacacaaatgctggcgcgacatgctgcctcttcgcgagcccgcttctgctgc	71093		
QY 244	tctcagcagaagattcacacaagcgtatccgaagggtgtgcgccagcgccgcccgcg	296		
Db 71092	aaccacacagcgacccacgcgcgcgcatggcccacgcgcgaccccgccgctcccg	71040		

```
misc_feature /clone.lib="CITB Mouse BAC
1.1060
misc_feature /note="assembly-fragment"
1161..2600
misc_feature /note="assembly-fragment"
2701..4058
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2001, 03:09:52 ; Search time 7860.64 Seconds  
(without alignments)  
115.238 Million cell updates/sec

Title: US-09-477-392-5  
Perfect score: 177  
Sequence: 1 cgatccgagggtcggccca.....tcgaagccccctcctc 177

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 255875100 residues  
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_ph: \*  
6: gb\_pl1: \*  
7: gb\_pl2: \*  
8: gb\_pr1: \*  
9: gb\_pr2: \*  
10: gb\_pr3: \*  
11: gb\_ro: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: em\_fun: \*  
15: em\_hum1: \*  
16: em\_hum2: \*  
17: em\_in: \*  
18: em\_om: \*  
19: em\_or: \*  
20: em\_ov: \*  
21: em\_pat: \*  
22: em\_ph: \*  
23: em\_pl: \*  
24: em\_ro: \*  
25: em\_sts: \*  
26: em\_sy: \*  
27: em\_un: \*  
28: em\_v1: \*  
29: gb\_htg1: \*  
30: gb\_htg2: \*  
31: gb\_in1: \*  
32: gb\_in2: \*  
33: em\_ba1: \*  
34: em\_ba2: \*  
35: em\_hum3: \*  
36: em\_hum4: \*  
37: gb\_pr4: \*  
38: gb\_htg3: \*  
39: gb\_htg4: \*  
40: gb\_htg5: \*  
41: gb\_htg6: \*  
42: gb\_htg7: \*  
43: em\_htg1: \*

44: em\_htg2: \*  
45: em\_htg3: \*  
46: em\_hum5: \*  
47: gb\_pl3: \*  
48: gb\_pr5: \*  
49: gb\_htg8: \*  
50: gb\_htg9: \*  
51: gb\_htg10: \*  
52: gb\_htg11: \*  
53: gb\_htg12: \*  
54: gb\_htg13: \*  
55: gb\_htg14: \*  
56: gb\_in3: \*  
57: gb\_htg15: \*  
58: gb\_htg16: \*  
59: gb\_htg17: \*  
60: em\_htg4: \*  
61: em\_htg5: \*  
62: em\_htg6: \*  
63: em\_htg7: \*  
64: em\_hum6: \*  
65: gb\_htg18: \*  
66: gb\_htg19: \*  
67: gb\_htg20: \*  
68: gb\_htg21: \*  
69: gb\_htg22: \*  
70: gb\_htg23: \*  
71: gb\_v11: \*  
72: gb\_v12: \*  
73: gb\_ba3: \*  
74: em\_htg8: \*  
75: em\_htg9: \*  
76: em\_htg10: \*  
77: gb\_pr6: \*  
78: gb\_pr7: \*  
79: gb\_sts1: \*  
80: gb\_sts2: \*  
81: gb\_pat1: \*  
82: gb\_pat2: \*  
83: em\_htg0: \*  
84: gb\_htg24: \*  
85: gb\_pr8: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query length	DB ID	Description
1	177	100.0	2878	37 AK025356	AK025356 Homo sapi
2	177	100.0	2878	77 HS245553	AJ245553 Homo sapi
3	177	100.0	2954	37 AF201303	AF201303 Homo sapi
4	177	100.0	132150	9 AC005586	AC005586 Homo sapi
5	52.4	29.6	166341	39 AC013381	AC013381 Homo sapi
6	52.4	29.6	244410	58 AC069214	AC069214 Homo sapi
7	52.2	29.5	63121	54 AC036107	AC036107 Homo sapi
8	52.2	29.5	131602	54 AC034263	AC034263 Homo sapi
9	50.2	28.1	78690	41 AC016541	AC016541 Homo sapi
10	49.8	28.1	8058	73 SC856	AC034263 Homo sapi
11	49.2	27.8	145540	48 AP001052	AL138852 Streptomy
12	49.2	27.8	340000	48 AP001052	AP001052 Homo sapi
13	49	27.7	65523	51 AC024266	AC024266 Homo sapi
14	48.8	27.6	72360	50 AC022984	AC022984 Homo sapi
15	48.6	27.5	68581	53 AC027375	AC027375 Homo sapi
16	48.4	27.3	28826	2 MTY041	AL021958 Mycobacte
17	48.2	27.2	135119	38 AC011578	AC011578 Homo sapi
18	47.8	27.0	63082	50 AC022663	AC022663 Homo sapi
19	47.6	26.9	93821	49 AC021596	AC021596 Homo sapi
20	47.6	26.9	151360	58 AC069132	AC069132 Homo sapi
21	47.2	26.7	340000	77 HS21C102	AL163302 Homo sapi



C	22	47	26.6	2028	47	SS132828	Ali32828 Spermatzo
C	23	47	26.6	61676	55	AC060800	Homo sapi
C	24	47	26.6	208936	38	AC010821	AC010821 Homo sapi
C	25	46.8	26.4	5186	73	SADNAIT2	X87940 S. aurantiac
C	26	46.8	26.4	40901	73	SCC24	Ali33003 Streptomy
C	27	46.8	26.4	159715	65	AC079369	AC079369 Mus muscu
C	28	46.4	26.2	371123	73	SC5F7	AL096872 Streptomy
C	29	46.4	26.2	59978	50	AC022815	AC022815 Homo sapi
C	30	46.4	26.2	109488	77	HS289J15	AL031774 Homo sapi
C	31	46.4	26.2	193782	65	AC079276	AC079276 Mus muscu
C	32	46.4	26.2	215218	59	AC076968	AC076968 Homo sapi
C	33	46.4	26.2	213661	66	ALI13825	ALI13825 Homo sapi
C	34	46.2	26.1	61676	55	AC060800	AC060800 Homo sapi
C	35	46.2	26.1	136551	55	AC048354	AC048354 Homo sapi
C	36	46.2	26.1	141952	67	ALI161774	ALI161774 Homo sapi
C	37	46.2	26.1	208279	39	AC013564	AC013564 Homo sapi
C	38	46	26.0	54298	65	AC080041	AC080041 Homo sapi
C	39	45.8	25.9	4910	78	HS963	X69910 H.sapiens p
C	40	45.8	25.9	12844	2	MBU34849	U34849 Mycobacteri
C	41	45.8	25.9	38500	2	MTCY39	Z74025 Mycobacteri
C	42	45.8	25.9	70408	66	AC083993	AC083993 Homo sapi
C	43	45.8	25.9	153559	52	AC025500	AC025500 Homo sapi
C	44	45.8	25.9	160326	59	AC073530	AC073530 Homo sapi
C	45	45.8	25.9	162976	53	AC026302	AC026302 Homo sapi

## ALIGNMENTS

BASE COUNT	560 a	974 c	826 g	518 t	100.0%; Score 177; DB 37; Length 2878; Best Local Similarity 100.0%; Pred. No. 1.4e-18; Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
ORIGIN	/note="Highly similar to Hs24553 Homo sapiens mRNA for Ap4 zinc finger protein"				
Query Match					
Db	1	cgatccgaagggtcggcccagcgcgcgcccccgccgggagagcccccagctgcacgcccgc	60		
Db	970	CGATCCGAGGGGATCGGCCACAGGCCGCCGCCGGGAGGACCCACGTCGACAGCGGCG	1029		
Qy	61	ccccgaggtcgcggcgccagagccaccccccggtgtaaccttgaaacccgcccagagaccg	120		
Db	1030	CCCCAGGAGTCCGGCGGCCGAGGCCACCCCGGGGTACTCTTAACCCGGCCACGAGGAGCGG	1089		
Qy	121	ccgcagagggccccgcacagacaccgcagagaccgagatgaagccccccctccctc	177		
Db	1090	CCGCCAGGGGCCCGCCGACAGAGCACCCCGCAGACCGGATGGAAGCCCCCCCCCTCCCTC	1146		

RESULT	1
AK025356	
LOCUS	
DEFINITION	AK025356 Homo sapiens cDNA: FLJ21703 fis, clone COL09895, highly similar to HS245553 Homo sapiens mRNA for AP4 zinc finger protein.
ACCESSION	AK025356
VERSION	AK025356.1 GI:10437856
KEYWORDS	oligo capping; ffs (full insert sequence).
SOURCE	Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL09895.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (sites) Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEO human cDNA sequencing project Unpublished (2000)
TITLE	2 (bases 1 to 2878) Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y. Direct Submission
JOURNAL	Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokawa-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cda@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, fax:81-3-5449-5416)
COMMENT	NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, Agency).  Location/Qualifiers 1..2878
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="COL09895" /clone_lib="COL" /tissue_type="colon" /note="Cloning vector pME18SFL3"
SOURCE	1..2878
misc.feature	

LOCUS	HSA245553	2878 bp	mRNA	PRI	18-AUG-1999
DEFINITION	Homo sapiens mRNA for AP4 zinc finger protein.				
ACCESSION	AJ245553				
VERSION	AJ245553.1	GI:5748564			
KEYWORDS	AP4 gene; AP4 protein; zinc finger protein.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 2878)				
JOURNAL	Dobner,T.G., Fischer,M. and Grottel,P.				
REFERENCE	Cloning of a novel zinc finger protein				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 2878)				
JOURNAL	Dobner,T.G.				
FEATURES	Direct Submission				
SOURCE	Submitted (17-AUG-1999) Dobner T.G., Institut fuer Medizinische				
	Mikrobiologie und Hygiene, Universitaet Regensburg,				
	Franz-Josef-Strauss-Allee 11, 93053 Regensburg, GERMANY				
	Location/Qualifiers				
	1..2878				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_line="HeLa"				
gene	106..1776				
	/gene="AP4"				
CDS	106..1776				
	/gene="AP4"				
	/note="zinc finger protein"				
	/codon_start=1				
	/product="AP4 protein"				
	/protein_id="CA853100.1"				
	/db_xref="GI:5748565"				
	/translation="MGPAQPRLLSGPQESFQTIGKSRGLRGOOGTSVAOSGAQAPEGH				
	AHRCARHRHPGVMALMLTIRCOQLPLEPCPGRRFAAPLAIHQVNAATPD				
	LGRACHICGSFGFWVALVLHLRAHSAKPRLIACPKEFRWRKOLRAHLCORCPA				
	PAAPRFICNGCGSFQAMODLVAKRVHALEAAKAVLGVRSGRPATVAPRGSGA				
	DADRPFCOACCGRKRHKPNLIARHVHGPERGPCGKRPNTKYLSHRITHT				
	GEPYRCKECGGRRFRHKPNLISHKTKRSBGSQAAPGSPDLPAGPOBSAEPTTP				
	AVPLKPAQEPGPAGPEHPQDPLEAPSLVSCDDGSRFLERFLAHQRHNGEPPFP				
	TCACEKGFKKTHLVASHSVSHSGEPFACEEGCRFSOGSHLAARDHADPDPFVC				
	PDGCAFRKHPLYLAHRIHTGEKPYCPDGCAFRKPSKNSLYSHRLHTGBSYACBDP				
	CDSFSKSNLTIRKSHINDRGCCALCGTFDEDERLLAHOKKHNV"				
BASE COUNT	528 a	973 c	839 g	538 t	
ORIGIN					
Query Match	100.0%; Score 177; DB 77; Length 2878;				

61 cccagagtcgcgcgcgagccaccgccgcgtacctctgaaccgcgccagagccg 120

Db 1.135 CCCGAGAGTCCGCGCGCCGAGCCACCCCGCGGTACCTGTGAACCGGCCGAGAGCCG 1.194

[illegible]

RESULT	4
AC005586/c	
LOCUS	
DEFINITION	AC005586 132150 bp DNA PRI 30-SEP-2000
	Homo sapiens PAC clone RPA-584D14 from 7q31-q35, complete sequence

RESULT	4		
AC005586/c			
LOCUS	AC005586	132150 bp	DNA
DEFINITION	Homo sapiens PAC clone	RP4-584D14	PRI
ACCESSION	AC005586		30-SEP-2000
VERSION	AC005586.2	GI:8468933	complete sequence
KEYWORDS	HTG.		
SOURCE	human.		

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1 (bases 1 to 132150)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	

REFERENCE	1 (bases 1 to 132150)
AUTHORS	Sulston, J.E. and Waterston, R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
REFERENCE	2 (bases 1 to 132150)

REFERENCE 2 (bases 1 to 132150)  
AUTHORS Stoneking, T., Ozersky, P., Woldmann, P. and Le, T.  
TITLE The Sequence of Homo sapiens PAC clone RP4-584D14  
JOURNAL Unpublished  
3 (bases 1 to 132150)  
REFERENCE

REFERENCE 3 (bases 1 to 132150)  
AUTHORS Waterston, K. H.  
TITLE Direct Submission  
SUBMITTED (01-SEP-1998) Genome Sequencing Center, Washington  
JOURNAL Submitted 444 Forest Park Parkway St Louis  
University School of Medicine

UNIVERSITY SCHOOL OF MEDICINE, 7444 LESTER BLVD, DC, DOWNS  
MO 63108, USA  
4 (bases 1 to 132150)  
Waterston, R. H.  
Direct Submission  
TITLE

**TITLE** Direct Submission  
**JOURNAL** Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis MO 63108, USA  
**E-ADDRESS** 1-800-368-123150

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

JOURNAL  
 Submitted (30-Sep-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jun 12, 2000 this sequence version replaced gi.3907529.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics -----

```

----- Summary Statistics -----
Center project name: H_DJ0584D14
-----
NOTICE: This sequence may not represent the entire insert of this

```

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats: all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping



Db 9738 CGATCCGAGGGGTGGCCCGACAGCCGCCGCCCGGCGGAGGCCGCCAGCTGCACAGCCGCG 9679

QY 61 ccccgagagtcggcgccgagccacaccccgcgctactctctgaacggcccgagcg 120  
|||||

Db 9678 CCCCGAGGAGTCCGCGCGCCGACGCCACCCCGGGGGTACTCTGAACCGCGCCGAGGAGCGG 9619

QY 121 cgcgcagggcgcccgccagagaccccgagagcccgatgaagccccctctcc 177  
|||||

Db 9618 CCGCAGGCGCCCGCCGACAGCACCCGCGAGAGCCGATCGAAGCCCGCCCTCTCC 9562

RESULT 5

AC013381

LOCUS AC013381

DEFINITION Homo sapiens chromosome 2 clone RP11-353M23 map 2, LOW-PASS

ACCESSION AC013381

VERSION AC013381.4 GI:9123907

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 166341)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Unpublished  
Homo sapiens chromosome 2, clone RP11-353M23  
2 (bases 1 to 166341)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baldwin, J., Barna, N., Beckert, R., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donnell, L., Doyle, M.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,  
Galagan, J., Gardy, S., Grant, G., Hasos, B., Hefford, A., Horton, L.,  
Howard, J. C., Johnson, R., Jones, C., Kamp, L., Karas, A., Klein, J.,  
Lewandzky, J., Lien, C., Locke, K., MacDonald, P., Margis, N.,  
McMan, P., McGuire, A., McKernan, K., McLaughlin, T., Melotim, J.,  
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6910757.  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html

COMMENT ----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: I3769  
Center clone name: 353\_M\_23

\* NOTE: This record contains 174 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 844: contig of 844 bp in length  
\* 845 944: gap of 100 bp  
\* 945 1799: contig of 855 bp in length  
\* 1800 1899: gap of 100 bp

*	1900	2775:	contly of 876 bp	in length
*	2776	2875:	gap of 100 bp	
*	2876	3771:	contly of 866 bp	in length
*	3772	3871:	gap of 100 bp	
*	3872	4761:	contly of 890 bp	in length
*	4762	4861:	gap of 100 bp	
*	4862	5756:	contly of 875 bp	in length
*	5737	5836:	gap of 100 bp	
*	5837	6640:	contly of 804 bp	in length
*	6641	6740:	gap of 100 bp	
*	6741	7617:	contly of 877 bp	in length
*	7618	7717:	gap of 100 bp	
*	7718	8586:	contly of 879 bp	in length
*	8397	8696:	gap of 100 bp	
*	8697	9555:	contly of 859 bp	in length
*	9556	9655:	gap of 100 bp	
*	9656	10517:	contly of 862 bp	in length
*	10518	10617:	gap of 100 bp	
*	10618	11478:	contly of 861 bp	in length
*	11479	11578:	gap of 100 bp	
*	11579	12420:	contly of 842 bp	in length
*	12421	12520:	gap of 100 bp	
*	12521	13422:	contly of 902 bp	in length
*	13423	13522:	gap of 100 bp	
*	13523	14362:	contly of 860 bp	in length
*	14383	14482:	gap of 100 bp	
*	14483	15347:	contly of 865 bp	in length
*	15348	15447:	gap of 100 bp	
*	15448	16303:	contly of 856 bp	in length
*	16304	16403:	gap of 100 bp	
*	16404	17266:	contly of 866 bp	in length
*	17270	17369:	gap of 100 bp	
*	17370	18239:	contly of 870 bp	in length
*	18240	18339:	gap of 100 bp	
*	18340	19192:	contly of 853 bp	in length
*	19193	19292:	gap of 100 bp	
*	19293	20160:	contly of 868 bp	in length
*	20161	20260:	gap of 100 bp	
*	20261	21120:	contly of 860 bp	in length
*	21121	21220:	gap of 100 bp	
*	21221	22076:	contly of 856 bp	in length
*	22077	22176:	gap of 100 bp	
*	22177	23040:	contly of 864 bp	in length
*	23041	23140:	gap of 100 bp	
*	23141	24000:	contly of 860 bp	in length
*	24001	24100:	gap of 100 bp	
*	24101	24966:	contly of 866 bp	in length
*	24967	25066:	gap of 100 bp	
*	25067	25954:	contly of 888 bp	in length
*	25955	26054:	gap of 100 bp	
*	26055	26916:	contly of 862 bp	in length
*	26917	27016:	gap of 100 bp	
*	27017	27891:	contly of 875 bp	in length
*	27892	27991:	gap of 100 bp	
*	27992	28859:	contly of 868 bp	in length
*	28860	28959:	gap of 100 bp	
*	28960	29827:	contly of 868 bp	in length
*	29828	29927:	gap of 100 bp	
*	29928	30828:	contly of 901 bp	in length
*	30829	30928:	gap of 100 bp	
*	30929	31742:	contly of 814 bp	in length
*	31743	31842:	gap of 100 bp	
*	31843	32699:	contly of 857 bp	in length
*	32700	32799:	gap of 100 bp	
*	32800	33662:	contly of 863 bp	in length
*	33663	33762:	gap of 100 bp	
*	33763	34638:	contly of 876 bp	in length
*	34639	34738:	gap of 100 bp	
*	34739	35603:	contly of 865 bp	in length
*	35604	35703:	gap of 100 bp	
*	35704	36578:	contly of 875 bp	in length
*	36579	36678:	gap of 100 bp	
*	36679	37519:	contly of 841 bp	in length

*	37520	37619:	gap of	100 bp		37520
*	37620	38483:	contig of	864 bp	in length	37620
*	38484	38583:	gap of	100 bp	in length	38484
*	38584	39427:	contig of	844 bp	in length	38584
*	39428	39527:	gap of	100 bp	in length	39428
*	39528	40387:	contig of	860 bp	in length	39528
*	40388	40487:	gap of	100 bp	in length	40388
*	40488	41342:	contig of	855 bp	in length	40488
*	41343	41442:	gap of	100 bp	in length	41343
*	41443	42291:	contig of	849 bp	in length	41443
*	42292	42391:	gap of	100 bp	in length	42292
*	42392	43246:	contig of	855 bp	in length	42392
*	43247	43346:	gap of	100 bp	in length	43247
*	43347	44206:	contig of	860 bp	in length	43347
*	44207	44306:	gap of	100 bp	in length	44207
*	44307	45145:	contig of	843 bp	in length	44307
*	45150	45249:	gap of	100 bp	in length	45150
*	45250	46137:	contig of	888 bp	in length	45250
*	46138	46237:	gap of	100 bp	in length	46138
*	46238	47122:	contig of	885 bp	in length	46238
*	47123	47222:	gap of	100 bp	in length	47123
*	47223	48148:	contig of	926 bp	in length	47223
*	48149	48248:	gap of	100 bp	in length	48149
*	48249	49112:	contig of	864 bp	in length	48249
*	49113	49212:	gap of	100 bp	in length	49113
*	49213	50075:	contig of	863 bp	in length	49213
*	50076	50175:	gap of	100 bp	in length	50076
*	50176	51027:	contig of	852 bp	in length	50176
*	51028	51127:	gap of	100 bp	in length	51028
*	51128	51982:	contig of	855 bp	in length	51128
*	51983	52082:	gap of	100 bp	in length	51983
*	52083	52941:	contig of	859 bp	in length	52083
*	52942	53041:	gap of	100 bp	in length	52942
*	53042	53907:	contig of	866 bp	in length	53042
*	53908	54007:	gap of	100 bp	in length	53908
*	54008	54858:	contig of	851 bp	in length	54008
*	54859	54958:	gap of	100 bp	in length	54859
*	54959	55820:	contig of	862 bp	in length	54959
*	55821	55920:	gap of	100 bp	in length	55821
*	55921	56793:	contig of	873 bp	in length	55921
*	56794	56893:	gap of	100 bp	in length	56794
*	56894	57786:	contig of	833 bp	in length	56894
*	57787	57886:	gap of	100 bp	in length	57787
*	57887	58511:	contig of	625 bp	in length	57887
*	58512	58611:	gap of	100 bp	in length	58512
*	58612	59483:	contig of	872 bp	in length	58612
*	59484	59583:	gap of	100 bp	in length	59484
*	59584	60433:	contig of	856 bp	in length	59584
*	60440	60539:	gap of	100 bp	in length	60440
*	60540	61400:	contig of	861 bp	in length	60540
*	61401	61500:	gap of	100 bp	in length	61401
*	61501	62357:	contig of	857 bp	in length	61501
*	62358	62457:	gap of	100 bp	in length	62358
*	62458	63325:	contig of	868 bp	in length	62458
*	63326	63425:	gap of	100 bp	in length	63326
*	63426	64283:	contig of	858 bp	in length	63426
*	64284	64383:	gap of	100 bp	in length	64284
*	64384	65260:	contig of	877 bp	in length	64384
*	65261	65360:	gap of	100 bp	in length	65261
*	65361	66214:	contig of	854 bp	in length	65361
*	66215	66314:	gap of	100 bp	in length	66215
*	66315	67195:	contig of	881 bp	in length	66315
*	67196	67295:	gap of	100 bp	in length	67196
*	67296	68164:	contig of	889 bp	in length	67296
*	68185	68284:	gap of	100 bp	in length	68185
*	68285	69182:	contig of	898 bp	in length	68285
*	69183	69282:	gap of	100		

Db 41045 CGGCCCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCGCGCCCCCCCCCGGCG 41104

Qy 74 cggcgcagagccacccgcgcgctactctctgaacccggccagagacgcgcgcgcacagggcc 133

Db 41105 GGGCGCGGGGCGCCCCCCCCNNCCCCCNCNCCGCGCCNCCGCGCGCGCGCGCGCGCCNCCN 41164

Qy 134 cgcgcagagcaccgcgcgcagagaccgatcgatgaagccccccctcc 175

Db 41165 CGCCCCGCGCGCCCCCGCGCGCCCCGCGCGCGCGCGCGCCNCC 41206

RESULT 6

AC068214/C

LOCUS

DEFINITION AC069214 244410 bp DNA HTG 25-JUL-2000

AC069214

AC069214

AC069214.5 GI:9438534

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 244410)

Muzny,D.M., Adams,C., Bailey,M., Barberie,J., Blankenburg,K., Bodola,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondeljewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswald,C., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Qulles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Taboi,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabwah,A., Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 244410)

Worley,K.C.

Direct Submission

Submitted (22-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 25, 2000 this sequence version replaced gi:8468712.

COMMENT

Query Match	29.6%	Score 52.4	DB 39	Length 166341
Best Local Similarity	54.9%	Pred. No. 1.1		
Matches	89	Conservative	0	Mismatches 73
				Indels 0
				Gaps 0

Quality coverage: 5.8x in Q20 bases; sum-of-contigs estimation

---

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hbsscc.bcm.tmc.edu/docs/enhbank\\_draft\\_data.html](http://www.hbsscc.bcm.tmc.edu/docs/enhbank_draft_data.html))

\* NOTE: This is a 'working draft' sequence. It currently

1	40616:	contig of 40616 bp in length
*	40617	gap of unknown length
*	40717	contig of 28735 bp in length
*	69452	69551: gap of unknown length
*	69552	98530: contig of 28879 bp in length
*	98531	98630: gap of unknown length
*	98631	125865: contig of 27234 bp in length
*	125865	125964: gap of unknown length
*	125965	149778: contig of 23814 bp in length
*	149778	149878: gap of unknown length
*	149779	149879: contig of 18274 bp in length
*	168153	168252: gap of unknown length
*	168253	168621: contig of 17965 bp in length
*	168621	186317: gap of unknown length
*	186317	195371: contig of 9054 bp in length
*	195372	195477: gap of unknown length
*	195477	205033: contig of 9562 bp in length
*	205034	205133: gap of unknown length
*	205134	213189: contig of 8056 bp in length
*	213190	213289: gap of unknown length
*	213290	216820: contig of 3531 bp in length
*	216821	216920: gap of unknown length
*	216921	219913: contig of 2993 bp in length
*	219914	220013: gap of unknown length
*	220014	221186: contig of 1173 bp in length
*	221187	221286: gap of unknown length
*	221287	223290: contig of 2004 bp in length
*	223291	223390: gap of unknown length
*	223391	225255: contig of 1869 bp in length
*	225260	225355: gap of unknown length
*	225360	227287: contig of 1928 bp in length
*	227288	227387: gap of unknown length
*	227388	228766: contig of 1361 bp in length
*	228766	228866: gap of unknown length
*	228869	230383: contig of 1515 bp in length
*	230384	230483: gap of unknown length
*	230484	231501: contig of 1018 bp in length
*	231502	231601: gap of unknown length
*	231602	232788: contig of 1187 bp in length
*	232789	232888: gap of unknown length
*	232889	234297: contig of 1409 bp in length
*	234298	234397: gap of unknown length
*	234398	235686: contig of 1289 bp in length
*	235687	235786: gap of unknown length
*	235787	237411: contig of 1625 bp in length
*	237412	237511: gap of unknown length
*	237512	238909: contig of 1398 bp in length
*	238910	239009: gap of unknown length
*	239010	240145: contig of 1136 bp in length
*	240146	240245: gap of unknown length
*	240246	241599: contig of 1354 bp in length
*	241600	241699: gap of unknown length
*	241700	242871: contig of 1172 bp in length
*	242872	242971: gap of unknown length
*	242972	244410: contig of 1439 bp in length

BASE COUNT ORIGIN	a	c	g	t	others
	69075	53746	52266	66581	2742

29.68; Score 52.4; DB 58; Length 244410;

RESULT	7
AC036107/c	
LOCUS	AC036107 63121 bp DNA
DEFINITION	Homo sapiens chromosome 17 clone RP11-31IK17 map 17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC036107
VERSION	AC036107.1 GI:7523741
KEYWORDS	HTG; HTGS_PHASED.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 63121)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens chromosome 17, clone RP11-311K17
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 63121)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.

TITLE	Direct Submission
JOURNAL	Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker:

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----Project Information-----  
Center project name: L2373  
Center clone name: 311\_K\_17

\* NOTE: This record contains 72 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely

```
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 802: contig of 802 bp in length
* 803 902: gap of 100 bp
* 903 1682: contig of 780 bp in length
* 1683 1782: gap of 100 bp
* 1783 2579: contig of 797 bp in length
* 2580 2679: gap of 100 bp
* 2680 3486: contig of 807 bp in length
* 3487 3586: gap of 100 bp
* 3587 4397: contig of 811 bp in length
* 4398 4497: gap of 100 bp
* 4498 5301: contig of 804 bp in length
* 5302 5401: gap of 100 bp
* 5402 6192: contig of 791 bp in length
* 6193 6292: gap of 100 bp
* 6293 7085: contig of 793 bp in length
* 7086 7185: gap of 100 bp
* 7186 7913: contig of 728 bp in length
* 7914 8013: gap of 100 bp
* 8014 8807: contig of 794 bp in length
* 8808 8907: gap of 100 bp
* 8908 9695: contig of 788 bp in length
* 9696 9795: gap of 100 bp
* 9796 10586: contig of 791 bp in length
* 10587 10686: gap of 100 bp
* 10687 11492: contig of 806 bp in length
* 11493 11592: gap of 100 bp
* 11593 12284: contig of 692 bp in length
* 12285 12384: gap of 100 bp
* 12385 13191: contig of 807 bp in length
* 13192 13291: gap of 100 bp
* 13292 14071: contig of 780 bp in length
* 14072 14171: gap of 100 bp
* 14172 14981: contig of 810 bp in length
* 14982 15081: gap of 100 bp
* 15082 15904: contig of 823 bp in length
* 15905 16004: gap of 100 bp
* 16005 16776: contig of 772 bp in length
* 16777 16876: gap of 100 bp
* 16877 17672: contig of 796 bp in length
* 17673 17772: gap of 100 bp
* 17773 18553: contig of 781 bp in length
* 18554 18653: gap of 100 bp
* 18654 19448: contig of 795 bp in length
* 19449 19548: gap of 100 bp
* 19549 20346: contig of 798 bp in length
* 20347 20446: gap of 100 bp
* 20447 21243: contig of 797 bp in length
* 21244 21343: gap of 100 bp
* 21344 22130: contig of 787 bp in length
* 22131 22230: gap of 100 bp
* 22231 23021: contig of 791 bp in length
* 23022 23121: gap of 100 bp
* 23122 23926: contig of 805 bp in length
* 23927 24026: gap of 100 bp
* 24027 24818: contig of 792 bp in length
* 24819 24918: gap of 100 bp
* 24919 25648: contig of 730 bp in length
* 25649 25748: gap of 100 bp
* 25749 26549: contig of 801 bp in length
* 26550 26649: gap of 100 bp
* 26650 27407: contig of 758 bp in length
* 27408 27507: gap of 100 bp
* 27508 28281: contig of 774 bp in length
* 28282 28381: gap of 100 bp
* 28382 29164: contig of 783 bp in length
* 29165 29264: gap of 100 bp
*
*
* 29265 29979: contig of 715 bp in length
* 29980 30079: gap of 100 bp
* 30080 30860: contig of 781 bp in length
* 30861 30960: gap of 100 bp
* 30961 31755: contig of 795 bp in length
* 31756 31855: gap of 100 bp
* 31856 32630: contig of 775 bp in length
* 32631 32730: gap of 100 bp
* 32731 33532: contig of 802 bp in length
* 33533 33632: gap of 100 bp
* 33633 34348: contig of 716 bp in length
* 34349 34448: gap of 100 bp
* 34449 35207: contig of 759 bp in length
* 35208 35307: gap of 100 bp
* 35308 36102: contig of 795 bp in length
* 36103 36202: gap of 100 bp
* 36203 37012: contig of 810 bp in length
* 37013 37112: gap of 100 bp
* 37113 37689: contig of 777 bp in length
* 37690 37989: gap of 100 bp
* 37990 38772: contig of 783 bp in length
* 38773 38872: gap of 100 bp
* 38873 39615: contig of 743 bp in length
* 39616 39715: gap of 100 bp
* 39716 40522: contig of 807 bp in length
* 40523 40622: gap of 100 bp
* 40623 41422: contig of 800 bp in length
* 41423 41522: gap of 100 bp
* 41523 42327: contig of 805 bp in length
* 42328 42427: gap of 100 bp
* 42428 43250: contig of 823 bp in length
* 43251 43350: gap of 100 bp
* 43351 44028: contig of 678 bp in length
* 44029 44128: gap of 100 bp
* 44129 44925: contig of 797 bp in length
* 44926 45025: gap of 100 bp
* 45026 45830: contig of 805 bp in length
* 45831 45930: gap of 100 bp
* 45931 46699: contig of 769 bp in length
* 46700 46799: gap of 100 bp
* 46800 47580: contig of 781 bp in length
* 47581 47680: gap of 100 bp
* 47681 48469: contig of 789 bp in length
* 48470 48569: gap of 100 bp
* 48570 49348: contig of 779 bp in length
* 49349 49448: gap of 100 bp
* 49449 50261: contig of 813 bp in length
* 50262 50361: gap of 100 bp
* 50362 51166: contig of 805 bp in length
* 51167 51266: gap of 100 bp
* 51267 52069: contig of 803 bp in length
* 52070 52169: gap of 100 bp
* 52170 52968: contig of 799 bp in length
* 52969 53068: gap of 100 bp
* 53069 53858: contig of 790 bp in length
* 53859 53958: gap of 100 bp
* 53959 54711: contig of 753 bp in length
* 54712 54811: gap of 100 bp
* 54812 55489: contig of 678 bp in length
* 55490 55589: gap of 100 bp
* 55590 56181: contig of 552 bp in length
* 56182 56281: gap of 100 bp
* 56282 57080: contig of 799 bp in length
* 57081 57180: gap of 100 bp
* 57181 57920: contig of 740 bp in length
* 57921 58020: gap of 100 bp
* 58021 58733: contig of 773 bp in length
* 58734 58833: gap of 100 bp
* 58834 59552: contig of 719 bp in length
* 59553 59652: gap of 100 bp
* 59653 60419: contig of 767 bp in length
* 60420 60519: gap of 100 bp
```





[illegible]

\* the record is updated, the accession number will  
\* be preserved.  
\* 1  
\* 839 938: contig of 838 bp in length  
\* 939 1827: contig of 889 bp in length  
\* 1828 1927: gap of 100 bp  
\* 1928 2811: contig of 884 bp in length  
\* 2812 2911: gap of 100 bp  
\* 2912 3744: contig of 833 bp in length  
\* 3745 3844: gap of 100 bp  
\* 3845 4762: contig of 918 bp in length  
\* 4763 4862: gap of 100 bp  
\* 4863 5733: contig of 871 bp in length  
\* 5734 5833: gap of 100 bp  
\* 5834 6688: contig of 865 bp in length  
\* 6689 6798: gap of 100 bp  
\* 6799 7671: contig of 873 bp in length  
\* 7672 7771: gap of 100 bp  
\* 7772 8633: contig of 862 bp in length  
\* 8634 8733: gap of 100 bp  
\* 8734 9578: contig of 845 bp in length  
\* 9579 9678: gap of 100 bp  
\* 9679 10560: contig of 882 bp in length  
\* 10561 10660: gap of 100 bp  
\* 10661 11550: contig of 890 bp in length  
\* 11551 11650: gap of 100 bp  
\* 11651 12510: contig of 860 bp in length  
\* 12511 12610: gap of 100 bp  
\* 12611 13446: contig of 836 bp in length  
\* 13447 13546: gap of 100 bp  
\* 13547 14426: contig of 880 bp in length  
\* 14427 14526: gap of 100 bp  
\* 14527 15417: contig of 891 bp in length  
\* 15418 15517: gap of 100 bp  
\* 15518 16378: contig of 861 bp in length  
\* 16379 16478: gap of 100 bp  
\* 16479 17337: contig of 859 bp in length  
\* 17338 17437: gap of 100 bp  
\* 17438 18328: contig of 891 bp in length  
\* 18329 18428: gap of 100 bp  
\* 18429 19298: contig of 870 bp in length  
\* 19299 19398: gap of 100 bp  
\* 19399 20220: contig of 822 bp in length  
\* 20221 20320: gap of 100 bp  
\* 20321 21214: contig of 894 bp in length  
\* 21215 21314: gap of 100 bp  
\* 21315 22202: contig of 888 bp in length  
\* 22203 22302: gap of 100 bp  
\* 22303 23160: contig of 858 bp in length  
\* 23161 23260: gap of 100 bp  
\* 23261 24084: contig of 824 bp in length  
\* 24085 24184: gap of 100 bp  
\* 24185 25053: contig of 869 bp in length  
\* 25054 25153: gap of 100 bp  
\* 25154 26032: contig of 879 bp in length  
\* 26033 26132: gap of 100 bp  
\* 26133 27010: contig of 878 bp in length  
\* 27011 27110: gap of 100 bp  
\* 27111 28001: contig of 891 bp in length  
\* 28002 28101: gap of 100 bp  
\* 28102 28999: contig of 898 bp in length  
\* 29000 29099: gap of 100 bp  
\* 29100 29958: contig of 859 bp in length  
\* 29959 30058: gap of 100 bp  
\* 30059 30948: contig of 890 bp in length  
\* 30949 31048: gap of 100 bp  
\* 31049 31936: contig of 888 bp in length  
\* 31937 32036: gap of 100 bp  
\* 32037 32919: contig of 883 bp in length  
\* 32920 33019: gap of 100 bp  
\* 33020 33869: contig of 850 bp in length  
\* 33870 33969: gap of 100 bp  
\* 33970 34850: contig of 881 bp in length  
  
\* 34851 34950: gap of 100 bp  
\* 34951 35844: contig of 894 bp in length  
\* 35845 35944: gap of 100 bp  
\* 35945 36851: contig of 907 bp in length  
\* 36852 36951: gap of 100 bp  
\* 36952 37848: contig of 897 bp in length  
\* 37849 37948: gap of 100 bp  
\* 37949 38744: contig of 796 bp in length  
\* 38745 38844: gap of 100 bp  
\* 38845 39745: contig of 901 bp in length  
\* 39746 39845: gap of 100 bp  
\* 39846 40711: contig of 866 bp in length  
\* 40712 40811: gap of 100 bp  
\* 40812 41688: contig of 877 bp in length  
\* 41689 41788: gap of 100 bp  
\* 41789 42644: contig of 856 bp in length  
\* 42645 42744: gap of 100 bp  
\* 42745 43612: contig of 868 bp in length  
\* 43613 43712: gap of 100 bp  
\* 43713 44559: contig of 847 bp in length  
\* 44560 44659: gap of 100 bp  
\* 44660 45529: contig of 870 bp in length  
\* 45530 45629: gap of 100 bp  
\* 45630 46515: contig of 886 bp in length  
\* 46516 46615: gap of 100 bp  
\* 46616 47501: contig of 886 bp in length  
\* 47502 47601: gap of 100 bp  
\* 47602 48484: contig of 883 bp in length  
\* 48485 48584: gap of 100 bp  
\* 48585 49438: contig of 854 bp in length  
\* 49439 49538: gap of 100 bp  
\* 49539 50413: contig of 875 bp in length  
\* 50414 50513: gap of 100 bp  
\* 50514 51390: contig of 877 bp in length  
\* 51391 51490: gap of 100 bp  
\* 51491 52359: contig of 869 bp in length  
\* 52360 52459: gap of 100 bp  
\* 52460 53320: contig of 861 bp in length  
\* 53321 53420: gap of 100 bp  
\* 53421 54316: contig of 896 bp in length  
\* 54317 54416: gap of 100 bp  
\* 54417 55274: contig of 858 bp in length  
\* 55275 55374: gap of 100 bp  
\* 55375 56245: contig of 871 bp in length  
\* 56246 56345: gap of 100 bp  
\* 56346 57209: contig of 864 bp in length  
\* 57210 57309: gap of 100 bp  
\* 57310 58212: contig of 903 bp in length  
\* 58213 58312: gap of 100 bp  
\* 58313 59192: contig of 880 bp in length  
\* 59193 59292: gap of 100 bp  
\* 59293 60169: contig of 877 bp in length  
\* 60170 60269: gap of 100 bp  
\* 60270 61097: contig of 828 bp in length  
\* 61098 61197: gap of 100 bp  
\* 61198 62100: contig of 903 bp in length  
\* 62101 62200: gap of 100 bp  
\* 62201 63095: contig of 895 bp in length  
\* 63096 63195: gap of 100 bp  
\* 63196 64079: contig of 884 bp in length  
\* 64080 64179: gap of 100 bp  
\* 64180 65024: contig of 845 bp in length  
\* 65025 65124: gap of 100 bp  
\* 65125 65999: contig of 875 bp in length  
\* 66000 66099: gap of 100 bp  
\* 66100 66973: contig of 874 bp in length  
\* 66974 67073: gap of 100 bp  
\* 67074 67946: contig of 873 bp in length  
\* 67947 68046: gap of 100 bp  
\* 68047 68942: contig of 896 bp in length  
\* 68943 69042: gap of 100 bp  
\* 69043 69930: contig of 888 bp in length  
\* 69931 70030: gap of 100 bp



```

gene
CDS
/protein_id="CAB72222.1"
/db_xref="GI:6912007"
/translacion="MVARSLSRANGSRRADGTFSCETMHAATGCTTTIRCAADY
RTAVSLVTRLATTVEDPAPRTADRAAPASERPSALMTSTLMSCLLASARE
PSTSDLPRLDRFEDRGHRNRTREDIDPEPTREHPRRLRLETGPELAPRRLDA
HPLPDLVAACEEHPDQPPRGADHIVRLRLDGAHVAVADGTGAVLWTRPDSGL
HPLVDSVALVLTLMALRRAGRLFTSRAGSGVREMAAAGRAL"
complement(2757..5078)
/gene="SCE56.04c"
complement(2757..5078)
/note="SCE56.04c, possible oxidoreductase, len: 773 aa;
similar to TR:09ZBN7 (EMBL:AL034492) Streptomyces
coelicolor putative oxidoreductase SC6C5.08, 828 aa; fasta
scores: opt: 1088 z-score: 1144.8 E(): 0; 42.2% identity
in 782 aa overlap and to SW:XDH.DKOPS (EMBL:M33977)
Drosophila pseudoobscura xanthine dehydrogenase (EC
1.1.1.204) Xdh, 1342 aa; fasta scores: opt: 517 z-score:
540.4 E(): 1.1e-22; 27.4% identity in 762 aa overlap.
Contains Pfam match to entry PF01315 Ald_Xan_dh_C,
Aldehyde oxidase and xanthine dehydrogenase, C terminus"
/codon_start=1
/transl_table=1
/product="putative oxidoreductase"
/protein_id="CAB72223.1"
/db_xref="GI:6912008"
/translacion="MSNDAATAEAGTAENATPVDPDIPHGIGASLPBPADARTEGTF
PPYADPAWAEGLMAVAIYRSPPRARIYSIDTTARAEMPGRAVTHEDVGSFLHGR
GGQADRPVFAEVAHNEGEPLAAVAADHAPRMAAAVAYEELDLPVNDPEAFAR
ELPHSDGRLIHIPRLGRGDPDPAAGEIVEGLIYRIGRDPDPAAGLADLPDRGVC
LYLSTDPHGGRDRIAAACGSLSPEDQKIVYTGVGATPDREDSOFOLPILALKTGC
PVKLATIEESLGHTHRHPTLRLRHADLEGRTVVEAOLILDAVADTSSDALA
AAVAAACGPVYPNAPAFIEGMVARTNPNPSGVREGAOVCAAEAMQDKVAKKLGID
PAEVLNRYNLATGVLDTPLGTQVTCPPAPVALLQAVRPLETLEKTPDEBMLPFGCP
EGAGSPVAGRVGSGVAGIWMHMGAEDESVTATVKGQAVATLCAAVETGGGFTT
LABRIYDITLVEYQVAVPDTPDPAGCPGCRGHRTHWGGAVRAAKMYRTOILDPL
AHQEMSTELLQIADGKITTSYDGLSTVYTTALDGKELMRTACGRHPPTLEDAQGG
DAFVEMATICALIRAVDVDTIEGSVRVELAAAOVGRYLNALQIATIEAGVTGQV
ALTENLRPRRLIRHPDLDTGLAALPDAIDRIYKIVLEERDVAVPGCAKAAVSAPV
ASPAVVASVRAAGRPNRRIPRPOAAVATGH"
complement(2943..4955)
/gene="SCE56.04c"
/note="Pfam match to entry PF01315 Ald_Xan_dh_C, Aldehyde
oxidase and xanthine dehydrogenase, C terminus, score
279..60, E-value 3..9e-80"
complement(5075..7396)
/gene="SCE56.05c"
complement(5075..7396)
/gene="SCE56.05c"
/note="SCE56.05c, unknown, len: 773 aa. High content in
proline, glycine and alanine. Contains Pfam matches to
entries PF00111 fer2, 2fe-2s iron-sulfur cluster binding
domains and PF01799 fer2.2, [2fe-2s] binding domain and
match to Prosite entry P800197 2fe-2s ferredoxins,
Iron-sulfur binding region signature"
/codon_start=1
/transl_table=1
/product="hypothetical protein SCE56.05c"
/protein_id="CAB72224.1"
/db_xref="GI:6912009"
/translacion="WTDQHQCDPHEDQHGQGTTPGSGRMDPLPGDYDDGATAFVKLP
EGGVYDALASDPSPLAAGHGVPPOLIVAPATTAAGDPARTGSMAAPSASAVDCTQ
OMOVYDADQDPLGAGYGYGPAQODPYGAARPVHNAQSGAPGADDEAFAPQPGCO
DESGAAQAQASRAADAAARQAQSGOVTQMSLPVADGDLPPESGEFTTSLAEQW
GCGATATLPDGAAPWATOPAPGCGWGEDAPRPGAPDDEERRAPASGHAYGATTHGG
BGHGADACHRRADDLSTGYPGHDSGADAGHPDDAAYGPHRGSPAPGAGYPVPAHNGC
STASPGHGSGHGPAAVSHGHDPAVSPGCRPAGCSTAHAGVDTQSAAGAAGAPGVN
RAPDAPSGSGGTATDGPGRGAEPAGPGRSGRAGHEERAKSRPEAAEGVATDPAAPTA
GERAPPEPAGSGEGQAPPEPSPAPAPVAGVADHRADSPDADPADYIDPDPDGG
KRTDAADSGADGDEAVTDGVPDPAQDQSPAPAPRODHPPLASVLYADLADRDVTTA
WGLATLDEPVSVDQRLARGAIVGCGFYCPGAAHTVHNILBENAPTELEARQALCGVL
CNCSTYRCVLEAVQVVAEREAHADADAGAPDDEEPGTPIDQAPGAGGVNAPAEFQO

```

```

RBS      pPGA0ATGAHDDPYPGDHDHGTHGMDGSHA"
        complement(5085..5089)

Query Match          28.1%; Score 49.8; DB 73; Length 8058;
Best Local Similarity 58.4%; Pred. No. 12;
Matches 87; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY       20 aggcgcccccgcggccgagcagccccagctgtccagccgcccccgagaagtccgycgcg 79
         ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db        6087 AGGGCGCTGTAAAGCCCCCGTGTACGAGCACCGGGCCGCCACCCCCTGGCGGGAGCGCGCCG 6028

QY       80 agccaccgccgcgtactctgtcaaacacggcccgaggagccgcgcgcgaagggcccccgccag 139
         ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db        6027 ACccccCGGGGGAACAACGGGGCAGCGCCCGAGAGAGAAGCGCGGCCGCAGTCCCCGCGCCG 5968

QY       140 agcaccgcagagccgatccgaagccccc 168
         | || | | |||| | |||| | || |
Db        5967 CCCCCAGGGCGCTCCCGCGCGAAGCGTC 5939


RESULT 11
LOCUS     AP001052/c                                PRI             14-JAN-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21, clone:KB22A5, MX1-D21S171
            region, complete sequence.
ACCESSION AP001052
VERSION    AP001052.1 GI:6693602
KEYWORDS   HTG.
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_11b:Keiold
            BAC library clone:KB22A5.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE    1 (bases 1 to 145540)
AUTHORS      Eukheria; Primates; Catarrhini; Homiidae; Homo.
TITLE        Shimizu,N., Kudoh,J. and Shibuya,K.
            Homo sapiens genomic DNA, chromosome 21, clone KB22A5, MX1-D21S171
REGION       region
NOTES        Published Only In Database (2000) In press
            2 (bases 1 to 145540)
            Shimizu,N., Kudoh,J. and Shibuya,K.
            Direct Submission
            Submitted (12-JAN-2000) to the DDBJ/EMBL/GenBank databases.
            NObuyoshi Shimomachi, Keio University, School of Medicine, Molecular
            Biology; 35 Shinanomachi, Shijuku-ku, Tokyo 160-8582, Japan
            (E-mail:nshimizu@med.keio.ac.jp, Tel:81-3-3351-2370,
            Fax:81-3-3351-2370)

FEATURES             Location/Qualifiers
     source           1..145540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="FLEB14-14"
                     /cell_type="pre-pro-B cell"
                     /chromosome="21"
                     /clone="KB22A5"
                     /clone_11b="Keio BAC library"
BASE COUNT          35992 a 36783 c 36979 g 35786 t
ORIGIN

```



[illegible]

```

RESULT 13
AC024266/c
LOCUS      AC024266      65523 bp      DNA      HTG      13-JUL-2000
DEFINITION Homo sapiens clone RP11-115F10, LOW-PASS SEQUENCE SAMPLING.
AC024266
VERSION    AC024266.1  GI:7107792
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 65523)
            Birren, B., Linton, L., Nussbaum, C. and Lander, E.
            Homo sapiens, clone RP11-115F10
            Unpublished
            2  (bases 1 to 65523)
            Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
            Anderson, S., Baldwin, J., Baran, N., Bedalov, F., Boguslavsky, L.,
            Bouckgeyser, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
            Choehel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
            D'Arlellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M.,
            Fenesator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
            Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
            Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
            Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
            Klein, J., Landers, T., Laroque, K., Lehotzky, J., Levine, R.,
            Lieu, C., Liu, G., Locke, K., MacDonald, P., Marguis, N., McCarthy, M.,
            McEwan, P., McGuirk, A., McKernan, K., McPheters, R., Meldrum, J.,
            Meneses, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
            Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T. M.,
            Peterson, K., Pierre, N., Pissini, C., Pollara, V., Raymond, C.,
            Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
            Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
            Sudirmanman, A., Talamas, J., Testaye, S., Theodore, J., Tittell, A.,
            Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
            Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
            Zody, M.
            Direct Submission
            Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            -----
            Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu

            -----
            Project Information
            Center project name: L1997
            Center clone name: 115_F_10

            -----
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 793: contig of 793 bp in length
            * 794 893: gap of 100 bp
            * 894 1687: contig of 794 bp in length
            * 1688 1787: gap of 100 bp
            * 1788 2572: contig of 785 bp in length
            * 2573 2672: gap of 100 bp

```



```
* 2673 3486: contig of 814 bp in length
* 3487 3586: gap of 100 bp
* 3587 4391: contig of 805 bp in length
* 4392 4491: gap of 100 bp
* 4492 5298: contig of 807 bp in length
* 5299 5398: gap of 100 bp
* 5399 6142: contig of 744 bp in length
* 6143 6242: gap of 100 bp
* 6243 7033: contig of 791 bp in length
* 7034 7133: gap of 100 bp
* 7134 7941: contig of 808 bp in length
* 7942 8041: gap of 100 bp
* 8042 8832: contig of 791 bp in length
* 8833 8932: gap of 100 bp
* 8933 9736: contig of 804 bp in length
* 9737 9836: gap of 100 bp
* 9837 10619: contig of 783 bp in length
* 10620 10719: gap of 100 bp
* 10720 11503: contig of 784 bp in length
* 11504 11603: gap of 100 bp
* 11604 12326: contig of 723 bp in length
* 12327 12426: gap of 100 bp
* 12427 13201: contig of 775 bp in length
* 13202 13301: gap of 100 bp
* 13302 14085: contig of 784 bp in length
* 14086 14185: gap of 100 bp
* 14186 14981: contig of 796 bp in length
* 14982 15081: gap of 100 bp
* 15082 15879: contig of 798 bp in length
* 15880 15979: gap of 100 bp
* 15980 16745: contig of 766 bp in length
* 16746 16845: gap of 100 bp
* 16846 17645: contig of 800 bp in length
* 17646 17745: gap of 100 bp
* 17746 18541: contig of 796 bp in length
* 18542 18641: gap of 100 bp
* 18642 19447: contig of 806 bp in length
* 19448 19547: gap of 100 bp
* 19548 20376: contig of 829 bp in length
* 20377 20476: gap of 100 bp
* 20477 21179: contig of 703 bp in length
* 21180 21279: gap of 100 bp
* 21280 22064: contig of 785 bp in length
* 22065 22164: gap of 100 bp
* 22165 22935: contig of 771 bp in length
* 22936 23035: gap of 100 bp
* 23036 23851: contig of 816 bp in length
* 23852 23951: gap of 100 bp
* 23952 24827: contig of 876 bp in length
* 24828 24927: gap of 100 bp
* 24928 25698: contig of 771 bp in length
* 25699 25798: gap of 100 bp
* 25799 26586: contig of 788 bp in length
* 26587 26686: gap of 100 bp
* 26687 27505: contig of 819 bp in length
* 27506 27605: gap of 100 bp
* 27606 28408: contig of 803 bp in length
* 28409 28508: gap of 100 bp
* 28509 29294: contig of 786 bp in length
* 29295 29394: gap of 100 bp
* 29395 30181: contig of 787 bp in length
* 30182 30281: gap of 100 bp
* 30282 31028: contig of 747 bp in length
* 31029 31128: gap of 100 bp
* 31129 31928: contig of 800 bp in length
* 31929 32028: gap of 100 bp
* 32029 32815: contig of 787 bp in length
* 32816 32915: gap of 100 bp
* 32916 33724: contig of 809 bp in length
* 33725 33824: gap of 100 bp
* 33825 34585: contig of 761 bp in length
* 34586 34685: gap of 100 bp
* 34686 35469: contig of 784 bp in length
```

```
* 35470 35569: gap of 100 bp
* 35570 36370: contig of 801 bp in length
* 36371 36470: gap of 100 bp
* 36471 37277: contig of 807 bp in length
* 37278 37377: gap of 100 bp
* 37378 38157: contig of 780 bp in length
* 38158 38257: gap of 100 bp
* 38258 39015: contig of 758 bp in length
* 39016 39115: gap of 100 bp
* 39116 39853: contig of 738 bp in length
* 39854 39953: gap of 100 bp
* 39954 40734: contig of 781 bp in length
* 40735 40834: gap of 100 bp
* 40835 41594: contig of 760 bp in length
* 41595 41694: gap of 100 bp
* 41695 42497: contig of 803 bp in length
* 42498 42597: gap of 100 bp
* 42598 43390: contig of 793 bp in length
* 43391 43490: gap of 100 bp
* 43491 44319: contig of 829 bp in length
* 44320 44419: gap of 100 bp
* 44420 45162: contig of 743 bp in length
* 45163 45262: gap of 100 bp
* 45263 46047: contig of 785 bp in length
* 46048 46147: gap of 100 bp
* 46148 46913: contig of 766 bp in length
* 46914 47013: gap of 100 bp
* 47014 47792: contig of 779 bp in length
* 47793 47892: gap of 100 bp
* 47893 48669: contig of 777 bp in length
* 48670 48769: gap of 100 bp
* 48770 49560: contig of 791 bp in length
* 49561 49660: gap of 100 bp
* 49661 50399: contig of 739 bp in length
* 50400 50499: gap of 100 bp
* 50500 51301: contig of 802 bp in length
* 51302 51401: gap of 100 bp
* 51402 52227: contig of 826 bp in length
* 52228 52327: gap of 100 bp
* 52328 53120: contig of 793 bp in length
* 53121 53220: gap of 100 bp
* 53221 53973: contig of 753 bp in length
* 53974 54073: gap of 100 bp
* 54074 54862: contig of 789 bp in length
* 54863 54962: gap of 100 bp
* 54963 55747: contig of 785 bp in length
* 55748 55847: gap of 100 bp
* 55848 56660: contig of 813 bp in length
* 56661 56760: gap of 100 bp
* 56761 57571: contig of 811 bp in length
* 57572 57671: gap of 100 bp
* 57672 58451: contig of 780 bp in length
* 58452 58551: gap of 100 bp
* 58552 59333: contig of 782 bp in length
* 59334 59433: gap of 100 bp
* 59434 60245: contig of 812 bp in length
* 60246 60345: gap of 100 bp
* 60346 61086: contig of 741 bp in length
* 61087 61186: gap of 100 bp
* 61187 61958: contig of 772 bp in length
```

Query Match 27.7%: Score 49; DB 51; Length 65523;  
Best Local Similarity 57.1%: Pred. No. 5.5;  
Matches 88; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
OY 24 cgcgccgagccggaggagccagctgcagcgagccgagagagtcgagcgccagagc 83
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 21163 ccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 21104
OY 84 caccgcgcgctactctcgaagagagagagagagagagagagagagagagagag 143
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 21103 cgcggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 21044
```

QY 144 ccgcaggaaccgatcgaaagccccccctccctc 177  
 ||| ||| | ||||| ||| |  
 Db 21043 ccccgccccccgccccccgcccccccccccc 21010

RESULT 14				
AC022984				
LOCUS	72360 bp	DNA	HTG	13-JUL-2000
DEFINITION	Homo sapiens chromosome 21 c clone RP11-565x18 map 21, LOW-PASS			



1594 1693: gap of 100 bp  
1694 2467: contig of 774 bp in length  
2468 2567: gap of 100 bp  
2568 3331: contig of 764 bp in length  
3332 3431: gap of 100 bp  
3432 4207: contig of 776 bp in length  
4208 4307: gap of 100 bp  
4308 5069: contig of 762 bp in length  
5070 5169: gap of 100 bp  
5170 5929: contig of 760 bp in length  
5930 6029: gap of 100 bp  
6030 6816: contig of 787 bp in length  
6817 6916: gap of 100 bp  
6917 7666: contig of 750 bp in length  
7667 7766: gap of 100 bp  
7767 8540: contig of 774 bp in length  
8541 8640: gap of 100 bp  
8641 9401: contig of 761 bp in length  
9402 9501: gap of 100 bp  
9502 10274: contig of 773 bp in length  
10275 10374: gap of 100 bp  
10375 11122: contig of 748 bp in length  
11123 11222: gap of 100 bp  
11223 11969: contig of 747 bp in length  
11970 12069: gap of 100 bp  
12070 12830: contig of 761 bp in length  
12831 12930: gap of 100 bp  
12931 13694: contig of 764 bp in length  
13695 13794: gap of 100 bp  
13795 14568: contig of 774 bp in length  
14569 14668: gap of 100 bp  
14669 15422: contig of 754 bp in length  
15423 15522: gap of 100 bp  
15523 16276: contig of 754 bp in length  
16277 16376: gap of 100 bp  
16377 17095: contig of 719 bp in length  
17096 17195: gap of 100 bp  
17196 17954: contig of 759 bp in length  
17955 18054: gap of 100 bp  
18055 18826: contig of 772 bp in length  
18827 18926: gap of 100 bp  
18927 19689: contig of 763 bp in length  
19690 19789: gap of 100 bp  
19790 20568: contig of 779 bp in length  
20569 20668: gap of 100 bp  
20669 21442: contig of 774 bp in length  
21443 21542: gap of 100 bp  
21543 22305: contig of 763 bp in length  
22306 22405: gap of 100 bp  
22406 23262: contig of 757 bp in length  
23263 23262: gap of 100 bp  
23263 24040: contig of 778 bp in length  
24041 24140: gap of 100 bp  
24141 24889: contig of 749 bp in length  
24890 24989: gap of 100 bp  
24990 25763: contig of 774 bp in length  
25764 25863: gap of 100 bp  
25864 26646: contig of 783 bp in length  
26647 26746: gap of 100 bp  
26747 27508: contig of 762 bp in length  
27509 27608: gap of 100 bp  
27609 28367: contig of 759 bp in length  
28368 28467: gap of 100 bp  
28468 29245: contig of 778 bp in length  
29246 29345: gap of 100 bp  
29346 30114: contig of 769 bp in length  
30115 30214: gap of 100 bp  
30215 30970: contig of 756 bp in length  
30971 31070: gap of 100 bp  
31071 31839: contig of 769 bp in length  
31840 31939: gap of 100 bp  
31940 32654: contig of 715 bp in length  
32655 32754: gap of 100 bp

32755 33517: contig of 763 bp in length  
33518 33617: gap of 100 bp  
33618 34379: contig of 762 bp in length  
34380 34479: gap of 100 bp  
34480 35241: contig of 762 bp in length  
35242 35341: gap of 100 bp  
35342 36077: contig of 736 bp in length  
36078 36177: gap of 100 bp  
36178 36927: contig of 750 bp in length  
36928 37027: gap of 100 bp  
37028 37777: contig of 750 bp in length  
37778 37877: gap of 100 bp  
37878 38623: contig of 746 bp in length  
38624 38723: gap of 100 bp  
38724 38493: contig of 770 bp in length  
38494 39593: gap of 100 bp  
39594 40360: contig of 767 bp in length  
40361 40460: gap of 100 bp  
40461 41219: contig of 759 bp in length  
41220 41319: gap of 100 bp  
41320 42090: contig of 771 bp in length  
42091 42190: gap of 100 bp  
42191 42953: contig of 763 bp in length  
42954 43053: gap of 100 bp  
43054 43829: contig of 776 bp in length  
43830 43929: gap of 100 bp  
43930 44697: contig of 768 bp in length  
44698 44797: gap of 100 bp  
44798 45557: contig of 760 bp in length  
45558 45657: gap of 100 bp  
45658 46395: contig of 738 bp in length  
46396 46495: gap of 100 bp  
46496 47248: contig of 753 bp in length  
47249 47348: gap of 100 bp  
47349 48118: contig of 770 bp in length  
48119 48218: gap of 100 bp  
48219 48976: contig of 758 bp in length  
48977 49076: gap of 100 bp  
49077 49834: contig of 758 bp in length  
49835 49934: gap of 100 bp  
49935 50692: contig of 758 bp in length  
50693 50792: gap of 100 bp  
50793 51517: contig of 725 bp in length  
51518 51617: gap of 100 bp  
51618 52351: contig of 734 bp in length  
52352 52451: gap of 100 bp  
52452 53215: contig of 764 bp in length  
53216 53315: gap of 100 bp  
53316 53968: contig of 654 bp in length  
53970 54069: gap of 100 bp  
54070 54828: contig of 759 bp in length  
54829 54928: gap of 100 bp  
54929 55690: contig of 762 bp in length  
55691 55790: gap of 100 bp  
55791 56554: contig of 764 bp in length  
56555 56654: gap of 100 bp  
56655 57423: contig of 769 bp in length  
57424 57523: gap of 100 bp  
57524 58295: contig of 772 bp in length  
58296 58395: gap of 100 bp  
58396 59176: contig of 781 bp in length  
59177 59276: gap of 100 bp

Query Match 27.5%; Score 48.6; DB 53; Length 68581;  
Best Local Similarity 53.8%; Pred. No. 6.2;  
Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 5 ccgaggggtcgcccgagcgcccgccgagggagcccgagcgagcccgcc 64  
Db 38437 CCAAGGCGCAACCAAGGGGCGCCGCCGACGCGCCCGCGCGCCGCC 38378  
QY 65 aggagtcgagcgagcgcccgcccgagcgagcgagcgagcgagcgagcg 124









```

/note="zinc finger protein"
/codon_start=1
/product="Ap4 protein"
/protein_id="CAB53100.1"
/db_xref="GI:5748365"
/translation="MGPAQPRLLSGPQSPQTLGKESRGLRQGTSTVAQSGAQBGR
AHCAHCRHHPGVALWLTTRCOARLPDPECCRRFRRHAFLLRHVHAATPD
LGFACHLGGSPFGWALVHLRAHSAKRPICPRCFRRFRKOLRAHLCORHPA
PAARPTCCNGCSFPAQMDQVLAHNRVHALEAALAEAAKYLGRSRGPRAVTAPRG
DAVDRPQCACCKRRFRRHNPILAHRRVHTGERPHOCPECGKFRFTNKPYLTSRRLHT
GKRPYCKECGRFRFRHNPILSHKILHKSSESAQAAPGSPQLAPGQESAABEP
ADPLKPAQEPPEPPAPPEHODPIEAPPSLYSCDDCRSEFLERFLRAHROHTGEPP
TCAECCKNFKGKTHLVASHRSVHSGERPFACCEGRRFSSHLAHRDHADAPRPYC
PQCKAFRRHKPYLAHRRIHTGKRPVPCDCKAFKOKSNLVSHRRIHTGERSVYACPD
CDRSEFSKSNLTHRSKSHIRIDGAFCCALICQGTDDERLLAHQKKHDV"

BASE COUNT      528 a      973 c      839 g      538 t
ORIGIN

Query Match      100.0%; Score 441; DB 77; Length 2878;
Best Local Similarity 100.0%; Pred. No. 5e-62;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtgagatgcgtcgacgcgccttcacgtgtgctgtgtgcaagcgttcgcgcac 60
|||
|||
|||
Db 754 GGTGAGATGCGCGTGCAGCCGCTTCAGTGTGCTGTGTGCAAGCGCTTCGGCAC 813
|||
|||
|||
QY 61 aagccaactgatacgtcaccgcgcgtgcaacagggcgagcgccaccagtgcccc 120
|||
|||
|||
Db 814 AAGCCCAACTGATCGCTCACCGCGGTCACACGGGCGCGCCGCCACACAGTGGCCC 873
|||
|||
|||
QY 121 gagtgcggagagcgcttaccataaagccctatctgaacttcgacggcgatccacac 180
|||
|||
|||
Db 874 GAGTGCGGAGAGCGCTTACCAATTAAGCCCTATCTGACTGTGCAACGGCGATCAACAC 933
|||
|||
|||
QY 181 ggcgagaagcctaccctgcaaaagatgcgcgcgcgttcgcgcacaaacccacactg 240
|||
|||
|||
Db 934 GCGGAGAAAGCCCTACCGGTGCAAAAGATGCGCGCGCTTCGGCAAAACCAACCTG 993
|||
|||
|||
QY 241 ctgtctacagcaagatcacaagagatccgagggatcgcccgacggcccgcccgccg 300
|||
|||
|||
Db 994 CTGTCTACAGCAAGATTACAAACGATTCGAGGGGTGCGGCCACGAGCCCGCGCGCGG 1053
|||
|||
|||
QY 301 gggagcccccaagtcgacggcgcccgagagatcgccggcgcccgcccgccggtta 360
|||
|||
|||
Db 1054 GGGAGCCCCCAGCTGCGACGGCGCCCGCAGAGATCGCGCGCCGACCCCGCGGTGTA 1113
|||
|||
|||
QY 361 cctctgaacccgagcccgagagcgcccgagggcgcccgcccgagagaccccgagacccg 420
|||
|||
|||
Db 1114 CCTCTGAACCGCGCCAGAGAGCCGCCGCCGCGCCGCGACAGACACCCCGCAGAGACCGG 1173
|||
|||
|||
QY 421 atcgaagccccccctccctc 441
|||
|||
|||
Db 1174 ATCGAAGCCCCCCCCCTCCCTC 1194
|||
|||
|||

RESULT 3
AF201303      2954 bp      mRNA      PRI      20-JAN-2000
DEFINITION   Homo sapiens dhfr oribeta-binding protein R1P60 mRNA, complete cds.
ACCESSION   AF201303
VERSION      AF201303.1 GI:6716713
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2954)
AUTHORS     Houchens,C.R., Montigny,W., Zeltser,L., Dailey,L., Gilbert,J.M. and
Heintz,N.H.
TITLE       The dhfr oribeta-binding protein R1P60 contains 15 zinc fingers:
DNA binding and looping by the central three fingers and an
associated proline-rich region

/note="zinc finger protein"
/codon_start=1
/product="Homo sapiens"
/protein_id="AF26712.1"
/db_xref="taxon:9606"
/cell_line="Hela"
/130..1833
/note="contains 15 C2H2-Kruppel-like zinc fingers"
/codon_start=1
/product="dhfr oribeta-binding protein R1P60"
/protein_id="AF26712.1"
/db_xref="GI:6716714"
/translation="MLERRCRGLPAMGLAQPRLLSGPQSPQTLGKESRGLRQGTST
VAQSGAQBGRHRAHCAHCRHHPGVALWLTTRCOARLPDPECCRRFRRHAFLLRHV
HAAATPDPLGFACHLGGSPFGWALVHLRAHSAKRPICPRCFRRFRKOLRAHLCOR
HPPAARPTCCNGCSFPAQMDQVLAHNRVHALEAALAEAAKYLGRSRGPRAVTAPRG
DAVDRPQCACCKRRFRRHNPILAHRRVHTGERPHOCPECGKFRFTNKPYLTSRRLHT
GKRPYCKECGRFRFRHNPILSHKILHKSSESAQAAPGSPQLAPGQESAABEP
ADPLKPAQEPPEPPAPPEHODPIEAPPSLYSCDDCRSEFLERFLRAHROHTGEPP
TCAECCKNFKGKTHLVASHRSVHSGERPFACCEGRRFSSHLAHRDHADAPRPYC
PQCKAFRRHKPYLAHRRIHTGKRPVPCDCKAFKOKSNLVSHRRIHTGERSVYACPD
CDRSEFSKSNLTHRSKSHIRIDGAFCCALICQGTDDERLLAHQKKHDV"

BASE COUNT      568 a      991 c      848 g      547 t
ORIGIN

Query Match      100.0%; Score 441; DB 37; Length 2954;
Best Local Similarity 100.0%; Pred. No. 4.9e-62;
Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtgagatgcgtcgacgcgccttcacgtgtgctgtgtgcaagcgttcgcgcac 60
|||
|||
|||
Db 811 GGTGAGATGCGCGTGCAGCCGCTTCAGTGTGCTGTGTGCAAGCGCTTCGGCAC 870
|||
|||
|||
QY 61 aagccaactgatacgtcaccgcgcgtgcaacagggcgagcgccaccagtgcccc 120
|||
|||
|||
Db 871 AAGCCCAACTGATCGCTCACCGCGGTGCAAAAGATGCGCGCGCTTCGGCAAAACCAACCTG 930
|||
|||
|||
QY 121 gagtgcggagagcgcttaccataaagccctatctgaacttcgacggcgatccacac 180
|||
|||
|||
Db 931 GAGTGCGGAGAGCGCTTACCAATTAAGCCCTATCTGACTGTGCAACGGCGATCAACAC 990
|||
|||
|||
QY 181 ggcgagaagcctaccctgcaaaagatgcgcgcgcgttcgcgcacaaacccacactg 240
|||
|||
|||
Db 991 GCGGAGAAAGCCCTACCGGTGCAAAAGATGCGCGCGCTTCGCGCACAAACCAACCTG 1050
|||
|||
|||
QY 241 ctgtctacagcaagatcacaagagatccgagggatcgcccgacggcccgcccgccg 300
|||
|||
|||
Db 1051 CTGTCTACAGCAAGATTACAAACGATTCGAGGGGTGCGGCCACGAGCCCGCGGTGTA 1110
|||
|||
|||
QY 301 gggagcccccaagtcgacggcgcccgagagatcgccggcgcccgcccgcccggtta 360
|||
|||
|||
Db 1111 GGGAGCCCCCAGCTGCGACGGCGCCCGCAGAGATCGCGCGCCGACCCCGCGGTGTA 1170
|||
|||
|||
QY 361 cctctgaacccgagcccgagagcgcccgagggcgcccgcccgagagaccccgagacccg 420
|||
|||
|||
Db 1171 CCTCTGAACCGCGCCAGAGAGCCGCCGCCGCGCGCCGCGACAGACACCCCGCAGAGACCGG 1230
|||
|||
|||
QY 421 atcgaagccccccctccctc 441
|||
|||
|||
Db 1231 ATCGAAGCCCCCCCCCTCCCTC 1251
|||
|||
|||

RESULT 4
AC005586/c    132150 bp    DNA      PRI      30-SEP-2000
LOCUS

```

DEFINITION Homo sapiens PAC clone RP4-584D14 from 7q31-q35, complete sequence.  
 AC005386  
 VERSION AC005386.2 GI:8468933  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 132150)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 132150)  
 AUTHORS Stoneking, T., Ozersky, P., Woldmann, P. and Le, T.  
 TITLE The sequence of Homo sapiens PAC clone RP4-584D14  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 132150)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-SEP-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 132150)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 132150)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jun 12, 2000 this sequence version replaced gi:3907529.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_DJ0584D14  
 -----  
 NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.  
 This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
 MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>  
 SOURCE INFORMATION:  
 This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

# FEATURES

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCYPAC2  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the right is RP5-820A21, 200 base pair overlap. Actual start of this clone is at base position 1 of RP4-584D14; actual end is at base position 131954 of RP4-584D14.  
 The sequence RP4-584D14 from base position 21655 to 22775 is GT rich. This region was sized with PCR from genomic DNA and the hindIII digest with band size 3643 real, and 3642 insilico. The sequence RP4-584D14 from base position 22628 to 22637 is represented by sequence derived by PCR from genomic DNA.  
 Location/Qualifiers  
 1..132150  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q31-q35"  
 /clone="RP4-584D14"  
 /clone\_1lb="RPCI-4"  
 230..315  
 /rpt\_family="MIR"  
 960..1288  
 /rpt\_family="MER4-group"  
 1360..1468  
 /rpt\_family="L2"  
 1562..1616  
 /rpt\_family="L2"  
 1733..2080  
 /note="similar to EST AW510685 (NID:97148763) hc89e08.x1"  
 2051..2114  
 /rpt\_family="L2"  
 2148..3224  
 /note="CpG island (%GC=70.7, o/e=0.90, #CpGs=107)"  
 2347..2426  
 /rpt\_family="GC-rich"  
 3141..3232  
 /rpt\_family="Alu"  
 3235..3368  
 /rpt\_family="Alu"  
 3795..3888  
 /rpt\_family="MIR"  
 4037..4199  
 /rpt\_family="Alu"  
 4308..4593  
 /rpt\_family="MER4-group"  
 4872..5309  
 /note="similar to EST A1827463 (NID:95448134) wf29f03.x1"  
 4882..4909  
 /rpt\_family="AT-rich"  
 5679..5773  
 /rpt\_family="MIR"  
 5947..5952  
 /note="match to EST AA482420 (NID:92210098) zt34f02.r1"  
 6148..6450  
 /rpt\_family="Alu"  
 6453..6690  
 /rpt\_family="MER4-group"  
 6695..7007  
 /rpt\_family="Alu"  
 7008..7117  
 /rpt\_family="MER4-group"  
 7180..7355  
 /rpt\_family="MER4-group"  
 7615  
 /note="match to EST A1144211 (NID:93666020) qb93g11.x1"  
 7883..8397  
 /note="match to EST W72943 (NID:91383235) zd54f12.r1"  
 7886..8357  
 /note="match to EST A1144211 (NID:93666020) qb93g11.x1"  
 7886..8562  
 misc\_feature

[illegible]

OY	241		ctcctcaacaagaattcaacaagcatccgaagggttcggcccacgaccgcgcccgccgc	300
Db	9762	CTGTCACAGCAGAATTTCAACAAGCCATTCCGAAGGGTGTGCACCAGGCCCCCCCGCCGC	9703	
OY	301	ggagagcccccaagtctgcacgcgcgcaccceaaaggatccgcgcgcgcagaccceacccgcgcgta	360	
Db	9702	GGGAGCCCCCAGTGTCGACGC GGCGCCCACAGAGATCTCGGGCGGCGAAGCCACCCGGCGGTa	9643	
OY	361	cctctgaacccgcgcgcacgaagcgccgc	420	
Db	9642	CCTGTGAACCGCGGCCACGAGGACCGCGCCGACGAGGCGCCCGCACAGAGCATCCGACGAGCACCCG	9583	
OY	421	atcgaagcccccccttccctc 441 		
Db	9582	ATCGAACGCCCCCTCTCTCTC 9562		
RESULT	5			
LOCUS	AC015887			
DEFINITION	Mus musculus clone C17-315E6, WORKING DRAFT SEQUENCE, 20 unordered pieces.			
ACCESSION	AC015887			
VERSION	AC015887.3			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E. 1 (bases 1 to 201458)			
JOURNAL	Mus musculus chromosome, clone C17-315E6			
REFERENCE	Unpublished 2 (bases 1 to 201458)			
AUTHORS	Baldwin,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Bateman,J., Barna,N., Beckerly,R., Boguslavskiy,L., Bouknighter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArtilano,K., Dewar,K., Domingo,M., Donegan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,G., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lien,C., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,P., Melidrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stefanovic,N., Sudramanlan,A., Talamas,J., Tesileye,S., Tirell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
COMMENT	All repeats were identified using RepeatMasker: Smitl,A.F.A. & Green, P. (1996-1997) <a href="http://f.f.g.washington.edu/RM/RepeatMasker.html">http://f.f.g.washington.edu/RM/RepeatMasker.html</a>			
	-- Genomes Center			
	Center: Whitehead Institute/ MIT Center for Genome Research			
	Center code: WIBR			
	Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a>			
	Contact: sequence_submissions@genome.wi.mit.edu			
	-- Project Information			
	Center project name: I742			
	Center clone name: 315_E_6			
	-- Summary Statistics			
	Sequencing vector: M13; M77815; 44% of reads			
	Sequencing vector: Plasmid; n/a; % 0-fraction of reads			
	55,555,555,555,556Chemistry: dye-primer-emershaw; 58% of reads			
	Chemistry: dye-terminator Big Dye; 42% of reads			
	Assembly program: Phrap; version 0.960731			
	Consensus quality: 192611 bases at least Q40			
	Consensus quality: 196150 bases at least Q30			
	Consensus quality: 197810 bases at least Q20			



This clone (DKFZp62K135) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp762K135"
/clone_1lb="762 (synonym: hme12). Vector pSPori1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"

```

BASE COUNT	ORIGIN
117 a	246 c
	252 g
	144 t

Query Match	32.3%;	Score 142.6;	DB 78;	Length 759;
Best Local Similarity	67.9%;	Pred. No. 4.9e-14;		
Matches 199; Conservative	0;	Mismatches 94;	Indels 0;	Gaps 0

Dy 8 atgcccgtcgaaacgcgcccttcacagtgtgacctgtgttgcaagcgcgtccgcgaagaagccca 67  
| | | | | | | | | | | | | | | | | | | | | |  
Db 709 acgcccccgatcgggccttctgctgtctccgactcggcgaaggcccttccgccacaaccct 655

QY 68 acttatacgtcaccgcgcgtgacacagggcgagcgcccccacagttccccgaattgcg 127

Db 649 ACCTGGCGGCACCGGGCATCCACACCGGGGAGAGCCCTACGTCTGCCCCGACTGCG 590

QY 128 ggaagcgcttaccataagccctatctgacttcgcaaccgycgcatacacaaccgycgaga 187

Db 589 GCAAGCCTTCAGCCAGAGTCCAACCTGGTGTCCGACCGGGCGCATCCACACGGGGGAGC 530

188 agccctaccgctgcgaagagtgcgccgcttcgycacaaacccaactgtgtctc 24

DB 329 GGCCTACGCTGTCCGACTTCGACCGCAGCTTCAGCCAGAATCCAACTTCAATCACC 4/

246 acagcaagallcaccagcgallccgaggggllcggcccaaggccgcccccgcccg 300

407 ACCCGCAAGACCCACACCACCGGACGGCCTTTCCTGCATGTGTTTGCCAA 411

## RESULT 7

LOCUS	AC073314	162570 bp	DNA	HTG	07-JUL-2000
-------	----------	-----------	-----	-----	-------------

SEQUENCE, 6 unordered pieces.

VERSION AC073314.2 GI:8954230

SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

REFERENCE	1
Watterson P H	(bases 1 to 162570)

**TITLE** The sequence of *Homo sapiens* clone

REFERENCE 2 (bases 1 to 1023/0)

AUTHORS Waterston, B. H.

**JOURNAL** Submitted (14-JUN-2000) Genome Sequencing Center, Washington

MO 63108, USA

AGGCGAGCGGCCCTCAAGTGCCCGAGTGCGAGAAGCTTCAGCGA

OY	71	tgatgtgcacacgcgcgtgcacacgagcgagcgagcccaacagtcgagtcgagga	130
Db	89940	TCACCAACCACATCGCCGCTGCATCGCGCGAGCGGCGCCGACCCCTGTCCCGAGTGGGCA	89999
OY	131	agcgcttaccataaagccctaactcgtacttcgacacgagcgacatccacacgagcgagaagc	190
Db	90000	AGAGCTTCATCCGCAAGCACCACTCTCTGGAGACACCGGGCGATCCACACAGCGGAGCGGC	90059
OY	191	ccctaacgcgtgcgaagatgtgcgcgcgcgtcttcgcgcacaaaccaactgtctgtctaca	250
Db	90060	CCTACCACTGCGCCCGAGTCCGGCAAGCGCGTTTACCCAGCAAGCATCACTGCTGGAGCAC	90119
OY	251	gcaagatccacaaagatattccgcagaggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	310
Db	90120	AGCGGCGGCACAGG---CGAGCGGCGCTTACCCCTGCAACGCACTGCGCAAGTGGTTCC	90176
OY	311	agctgcagacgc	370
Db	90177	-GCTCAAGCAATGCTCTCAATGACACCTGCGGACCCACACAGGGGAGTGAACGGCGCGC	90235
OY	371	cggccacagagc	430
Db	90236	CCGCGCGCGCGCCCGCGGCAGGTGCGCGGGCGTGCCTCCCTCTCGGACACCGCCAGGCC	90295
OY	431	cccc 435	
Db	90296	GAGCC 90300	

REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC004877/C	AC004877	Homo sapiens PAC clone RPA-751H13 from 7q35-qter, complete sequence.	AC004877	AC004877.1	GI:3638954	HTG.	human.
REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 128361)	Leonard, S., Graves, T., and Stromwater, C.	The sequence of Homo sapiens PAC clone RPA-751H13	Unpublished	2 (bases 1 to 128361)	Waterston, R.H.	Direct Submission	Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 128361)	Waterston, R.	Direct Submission	Submitted (19-SEP-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	4 (bases 1 to 128361)	Waterston, R.	Direct Submission	Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 19, 1998 this sequence version replaced gi:3213120.							

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics -----  
Center project name: H\_DD0751H13

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CNHR/>, send

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-811N16, 200 bp overlap. Actual start of this clone is at base position 1 of RP4-751H13; actual end is at 128361 of RP4-751H13.

## FEATURES

**Source**

repeat_region	12768..12890	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q35-qter" /clone="RP4-751H13" /clone_lib="unknown" 20..155
repeat_region	860..941	/rpt_family="Alu"
repeat_region	2351..2885	/rpt_family="MERL_type?"
repeat_region	2924..3246	/rpt_family="Retroviral"
repeat_region	3300..3727	/rpt_family="Alu"
repeat_region	4146..4287	/rpt_family="Retroviral"
repeat_region	5253..5728	/rpt_family="L2"
repeat_region	5933..6243	/rpt_family="MaLR"
repeat_region	7787..7870	/rpt_family="Alu"
repeat_region	7798..8145	/rpt_family="MIR"
repeat_region	8259..8311	/note="match to EST AA084720 (NID:g1626811) zn06f02.s1"
repeat_region	9100..9426	/rpt_family="L2"
repeat_region	9773..10086	/note="match to EST AA080867 (NID:g1623560) zn06f02.r1"
repeat_region	9947..10370	/rpt_family="Alu"
repeat_region	9947..10370	/note="match to EST AA828832 (NID:g2901931) od80c03.s1"
repeat_region	10066..10388	/note="match to EST AA828832 (NID:g2901931) od80c03.s1"
repeat_region	10297..10734	/note="match to EST AA857387 (NID:g2945689) od83h10.s1"
repeat_region	11605..12220	/rpt_family="MaLR"
repeat_region	12222..12615	/rpt_family="L1"
repeat_region	12617..12755	/rpt_family="Retroviral"
repeat_region	12768..12890	/rpt_family="L1"

```

repeat_region      /rpt_family="Alu"
                   23679..23794
repeat_region      /rpt_family="MIR"
                   23773..23846
repeat_region      /rpt_family="L2"
                   24894..24982
repeat_region      /rpt_family="L2"
                   24963..25271
repeat_region      /rpt_family="Alu"
                   25272..25616
misc_feature        /rpt_family="L2"
                   26933..29260
                   /note="CpG_island (%GC=73.8, o/e=0.90, #CpGs=195)"

Query Match      28.7%; Score 126.6; DB 8; Length 128361;
Best Local Similarity 67.8%; Pred. No. 1.9e-12;
Matches 177; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 11 cccgtgacccgccccttccagtgtgccttcttgtgcaagcgtctccggcacaagcccaact 70
Db 21081 CGGGGAGACGGGCCCTACCCGTGCGGGAGTGCGGCAAGGCGCTTCAGCCAGGAAGATCCACC 21022

OY 71 tgaatgcacacgcgcgtgtcacacggcgagcgcccccaacagtgccccgagtcggga 130
Db 21021 TGGGCTCCGACCAAAAGACCCACACCGCGGACGGCCCTTCCCTCGACAGATCGGACA 20962

OY 131 agcgcttaccataaagccatactgactctgcacccggcgcatccacacccgagcgaagc 190
Db 20961 AGCGCTTTCGGAGAAAGACGACTTGATTGGCGCACACGGGCAATCCATACAGGCGGAGAGCG 20902

OY 191 cctaccggtgcacaagagtgtgcgcgcgcttcgcgcacaaaccaactgtctctaca 250
Db 20901 CCTACCACTGGCGCACAGTGCAGCGACGACGCTTCACGCACACAGCAGCATGTTGTCGGGACCC 20842

OY 251 gcaagatccacaagcgtatcg 271
Db 20841 AAAGGCTGCACCGACGGCGG 20821

RESULT 9
AC013273/c
LOCUS Homo sapiens chromosome 7 clone RP11-428D5, WORKING DRAFT SEQUENCE.
DEFINITION 10 unorderd pieces.
AC013273
AC013273.5 GI:9211421
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 210923)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210923)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
COMMENT On Jul 15, 2000 this sequence version replaced gi:8990985.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Project name: H.NH0428D05
Center project name: Summary Statistics
-----
Sequencing vector: M13; 74#
Sequencing vector: plasmid; 26#
Chemistry: Dye-Primer ET; 74% of reads

```

*	1	2429: contig of 2429 bp in length
*	2430	2529: gap of unknown length
*	2530	4617: contig of 2088 bp in length
*	4618	4717: gap of unknown length
*	4718	1807: contig of 7090 bp in length
*	11808	11907: gap of unknown length
*	11908	20464: contig of 8557 bp in length
*	20465	20564: gap of unknown length
*	20565	29268: contig of 8734 bp in length
*	29269	29398: gap of unknown length
*	29399	4700: contig of 16302 bp in length
*	45701	45800: gap of unknown length
*	45801	63662: contig of 17862 bp in length
*	63663	63762: gap of unknown length
*	63763	104507: contig of 40745 bp in length
*	104508	104607: gap of unknown length
*	104608	150643: contig of 46036 bp in length
*	150644	150743: gap of unknown length
*	150744	210923: contig of 60180 bp in length

Query Match	28.7%;	Score 126.6;	DB 39;	Length 210923;
Best Local Similarity	67.8%;	Pred. No. 1.6e-12;		

RESULT	10			
LOCUS	MMZFP29	2094 bp	mRNA	13-APR-1992
DEFINITION	Musculus zfp-29 gene for zinc finger protein.			
ACCESSION	X55126			
VERSION	X55126.1	GI:55470		
KEYWORDS	spermatogenesis; zfp-29 gene; zinc-finger.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
AUTHORS	Rodentia; Sciurognath; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 2094)			
JOURNAL	Denny, P.			
	Direct Submission			
	Submitted (03-NOV-1990) Denny P., Institute of Cancer Research,			
	Chester Beatty Laboratories, 237 Fulham Road, LONDON, SW3 6JB, UK			
REFERENCE	2 (bases 1 to 2094)			
AUTHORS	Denny, P. and Ashworth, A.			
TITLE	A zinc finger protein-encoding gene expressed in the post-meiotidic			
JOURNAL	Phase of spermatogenesis			
MEDLINE	Gene 106 (2), 221-227 (1991)			
FEATURES	92039080			
source	Location/Qualifiers			
	1..2094			

gene 104. .1948  
/gene="%f-29"  
104. .1948  
/gene="%f-29"  
/odon\_start=1  
/protein\_id="CA38920.1"  
/db\_xref="GI:35471"  
/db\_xref="SWISS-PROT:Q07230"  
/translation="MAAEVPAVSTPLSLVGVPOEDEDAEFTVMILEDAMVQEAVALV  
QIEDGESPPFQDAGKSGPOEDAEAGGQALVYFRELCEKRWLENEVHTKCOMITLVLT  
REIQMLQENHPRESSEEAVALVEDLTQTFRISDELTISEKNEKNEEMKESVGHGFMF  
NISGSGEGQOQSDSDPERDGGSGGAGHAPGEDRVVPSEGGREVCOLIGLQSTYGLG  
EKPYPCPGCGKSTFKSRKSLILTHERHTGEXKYYKQDECKSPDSKNSRHOTHTGKRG  
PYKDEGCKGSRFSRANLITTHORIHGEGPQCAAGKGSFSRPLVLAHORHTGEXKRY  
SCPEGCKSGNRSSNLTHOHTHTGEXKRYACCEGSEPSYNSLTHORIHGEXKRYK  
TECCGKSEFQSSALITHRTHTGEXKRYQCGEGCKNFSRNSNLATIRHLVEXPKCEL  
CGKTSQSSSLVLAHQHTHTGEXKRYTECLTCGSEFSWMSLTHOHTHTGEXKRYCGDGG  
KGFORSQSLVLAHQHTHTGEXKRYKCLMGKGSFSRSLVMVHQRAHLGKPYKPCPGCKG  
FSWMSVLLIHHORIHGEXKRYKSCPEGCKGFSNMSNLTTHOHTHTGEXKRY"  
752. .1945





TITLE	REFERENCE	FEATURES
JOURNAL	Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System	Source
AUTHORS	Unpublished 3 (bases 1 to 154218) Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meinke,L., Longmire,J., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R., McWhirry,K., Han,C. and Deaven,L.	Location/Qualifiers 1. .154218 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /clone="RPCT-11_127120" complement(41. .350) /rpt_family="Alu" 575. .882 /rpt_family="Alu" 1414. .1712 /rpt_family="Alu" 1595. .1862 /rpt_family="Alu" 1871. .2190 /rpt_family="Alu" 2176. .2372 /rpt_family="L1" 2498. .2795 /rpt_family="Alu" 3187. .3504 /rpt_family="Alu" complement(3571. .3890) /rpt_family="Alu" complement(3859. .4163) /rpt_family="Alu" 4198. .4503 /rpt_family="Alu" 4826. .5155 /rpt_family="Alu" 5749. .6045 /rpt_family="Alu" complement(6848. .6983) /note="GRAIL 2 excellent exon, frame 1" 7473. .7776 /rpt_family="Alu" 7931. .8385 /rpt_family="Alu" 8808. .9139 /rpt_family="Alu" 9394. .9682 /rpt_family="Alu" 9515. .9812 /rpt_family="Alu" 9702. .9987 /rpt_family="Alu" 9965. .10332 /rpt_family="Alu" complement(10403. .10835) /rpt_family="Alu" complement(10801. .10977) /rpt_family="Alu" 11580. .11780 /rpt_family="MER42" 11803. .12101 /rpt_family="Alu" 12282. .12576 /rpt_family="Alu" 12643. .13073 /rpt_family="Alu"
TITLE	Direct Submission	
JOURNAL	Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA	

repeat_region	12923..13237	/rpt_family="Alu"	
repeat_region	13420..13725	/rpt_family="Alu"	
repeat_region	13761..14148	/rpt_family="Alu"	
repeat_region	14072..14371	/rpt_family="Alu"	
repeat_region	14401..14696	/rpt_family="Alu"	
repeat_region	15061..15366	/rpt_family="Alu"	
repeat_region	15383..15706	/rpt_family="Alu"	
repeat_region	16083..16185	/rpt_family="Alu"	
repeat_region	16192..16347	/rpt_family="Alu"	
repeat_region	16559..16631	/rpt_family="Alu"	
repeat_region	16820..18718	/rpt_family="Alu"	
repeat_region	18518..18917	/rpt_family="Alu"	
repeat_region	18754..19036	/rpt_family="Alu"	
misc_feature	19456..19559	/note="GRAL2 excellent exon, frame 1"	
repeat_region	20239..20397	/rpt_family="MER20"	
repeat_region	20808..21104	/rpt_family="Alu"	
repeat_region	22116..22443	/rpt_family="Alu"	
misc_feature	22789..22948	/note="GRAL2 excellent exon, frame 1"	
repeat_region	23346..23646	/rpt_family="Alu"	
repeat_region	24046..24341	/rpt_family="Alu"	
repeat_region	25205..25461	/rpt_family="Alu"	
repeat_region	25605..25662	/rpt_family="Alu"	
repeat_region	25914..25981	/rpt_family="Alu"	
repeat_region	26044..26355	/rpt_family="Alu"	
repeat_region	28437..28732	/rpt_family="Alu"	
misc_feature	30597..30772	/note="GRAL2 excellent exon, frame 2"	
repeat_region	31300..31655	/rpt_family="Alu"	
repeat_region	32059..32124	/rpt_family="Alu"	
repeat_region	32102..32492	/rpt_family="Alu"	
repeat_region	32507..32639	/rpt_family="Alu"	
repeat_region	33649..34040	/rpt_family="Alu"	
repeat_region	34202..34278	/rpt_family="LFR7B"	
repeat_region	34279..34556	/rpt_family="Alu"	
repeat_region	35017..35495	/rpt_family="Alu"	
repeat_region	35639..36991	/rpt_family="Alu"	
misc_feature	37667..37808	/note="GRAL2 excellent exon, frame 0"	

Query Match	27.98;	Score 123.2;	DB 10;	Length 154218;
Best Local Similarity	60.28;	Pred. No. 6.3e-12;		
Matches 242;	Conservative	0;	Mismatches 153;	Indels 7; Gaps 2

OY	8	atgcgctcgacagccccccttcacagtgtgcctgttggcaagaagcgtccgcgaagcca	67
Db	45624	ATTGGGCGTGACAAAGCCGTACACTCTGCCCGAATGTGGCAAAAGGCTTCACGAAGAAGTCCC	455655
OY	68	actgatgcgtcacaccgcgcgtgtgcaacagcyggaagcccccaagatgtgccagattgcg	127
Db	45564	ACTTAGCACAAGCACACAGCGACMACACAGCGGGGAGCGGCCTTAAGAATGTCCTGTTGG	455050
Db	128	ggaagcgctttaccaatagaaccttatctcgattgcgaecygcgcaltccaacacygagaa	187
Db	45504	GGAAGGCGTTTAGCAGCAGCGCTCCAACTTCAGAGCACACCAGAGGGGTGCACACAGCGCGAGA	454445
OY	188	aggccctaccgctgcaagaagatgtagccgcgcgcgtcttcggacaacaaccaaccatgcgtcttc	247
Db	45444	AGCCCTTACCCTGGCCCCGAGTGTGGGAAGCGCTTCAGCCAGAGCTCCAGCCTGTGTATCC	453855
OY	248	aacgaagaattcacaaagcatcgcgaagtgctgcgccaaagccgcgcccgcccgaggaagcc	307
Db	45384	AACCGAAGNACAC-----AGCGGGAGACGGCCCATCTCCTACACCAGTGGGGGAAGG	453330
OY	308	c ccagctgcagccgagcccccaagagatcgcgcgcgcgaagcccaaccccgycgtactcttga	367
Db	45329	CTTCAACAACAGCTGTGCACTTCAGAGGCCACACCGCGGAGCAGACAGGAGNAAGCCCTTA	452707
OY	368	aacgcgccccagagagcgcgcgcgcgaagygccccgcgaagagacc	409
Db	45269	CACCTGCCCC--GGCTGTGGCCGGGGCTTCCGCGCGGCGACC	452230

RESULT	13		
LOCUS	HSJ59G15/c		
DEFINITION	HSJ59G15 138278 bp	DNA	HTG
	Homo sapiens chromosome 1 clone RP4-59G15 map p12-13.2, ***		06-OCT-2000

```

ACCESSION      SEQUENCING IN PROGRESS ***, 6 unordered pieces.
VERSION        ALI09966
KEYWORDS       HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eulhetaia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      Donnelly, S.
AUTHORS        Submitted (04-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
TITLE          CB10 USA, UK. E-mail enquiries: humquerry@sanger.ac.uk
JOURNAL        On Aug 14, 2000 this sequence version replaced gl:9588436.
COMMENT        -----
                Center: Sanger Centre
                Genome Center
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                Project Information
                Center project name: dj599G15
                -----
                Summary Statistics
                Assembly program: XGAP4; version 4.5
                Sequencing vector: M13; M77815; 25% of reads
                Sequencing vector: plasmid; L08752; 74% of reads
                Chemistry: Dye-terminator ET-amersham; 0% of reads
                Chemistry: Dye-terminator ABI; 1% of reads
                Chemistry: Dye-terminator Big Dye; 94% of reads
                Chemistry: Dye-primer-amersham; 2% of reads
                Chemistry: Dye-primer Big Dye; 1% of reads
                Consensus quality: 136114 bases at least Q40
                Consensus quality: 137176 bases at least Q30
                Consensus quality: 137544 bases at least Q20
                Insert size: 137778; sum-of-contigs
                Insert size: 126614; 2.3% error; agarose-fp
                Quality coverage: 6.30x in Q20 bases; sum-of-contigs quality
                coverage: 8.38x in Q20 bases; agarose-fp
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 6 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                * 8020: contig of 8020 bp in length
                * 8021 8120: gap of 100 bp
                * 8121 15996: contig of 7876 bp in length
                * 15997 16096: gap of 100 bp
                * 16097 19055: contig of 2959 bp in length
                * 19056 19155: gap of 100 bp
                * 19156 22087: contig of 2932 bp in length
                * 22088 22187: gap of 100 bp
                * 22188 29998: contig of 7811 bp in length
                * 29999 30098: gap of 100 bp
                * 30099 138278: contig of 108180 bp in length.
                *
                Location/Qualifiers
                1. 138278
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="1"
                /map="p12-13.2"
                /clone="RP4-599G15"
                /clone_id="RPCT-4"
                1. 8020
                /note="assembly_fragment:01125
                fragment_chain:1"
                8121. 15996
                /note="assembly_fragment:02016
                fragment_chain:1"
                16097. 19055

```



REFERENCE 1 (bases 1 to 1824)  
AUTHORS Jansa, P. and Grummt, I.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-1997) Molecular Biology of the Cell II, German  
Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg, B.-W.  
D-69120, Germany

FEATURES  
source location/Qualifiers  
1..1824

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="WI-38"  
/tissue\_type="lung"  
/cell\_type="fibroblasts"  
/note="isolated by yeast two-hybrid screen"  
CDS  
1..1158  
/note="TTP20; Transcription Termination Factor I  
Interacting Peptide 20"  
/product="TTF-I interacting peptide 20"  
/protein\_id="AAB58413.1"  
/db\_xref="GI:2145060"

/translation="DPDMELREAFNSRPLDSEDOQMAHSHENPVILIMDDDVPSWP  
PTKLSPPOSAPRPPRPPAPYICNECGKSFHWSKLTROHTHTGERPNACADC  
GKTFSSOHLVORIRHTGEKPYACLECGKRFSSNLMOHRIHTEKPYTCPCGR  
SFQSKSLAKRRSHSGLPKPCVPCGRCGFSOPKSLARHLRLHPELSGPVAAKYLAA  
SVRAKGEPAVADGEIAPVGDGEGIIYVGAPEGCAAAAAMAGAKAAGPRSR  
APAPKPYVECECGKSGHAGLLAHORAGHDGLGAGGEPRAHICVECGEGFYQGA  
LRHKRIHAGAPSVSSCGSTYRAGGEDDDDEAAGRCPECRAGEGR"

BASE COUNT 307 a 611 c 655 g 251 t  
ORIGIN

Query Match 26.8%; Score 118; DB 10; Length 1824;  
Best Local Similarity 55.7%; Pred. No. 2.9e-10;  
Matches 226; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```

QY 11 ccgtcgacccgccccttcagtgctgtgtgcaagcgcttcgagcagaagcccaact 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 CCGGCGAGAGGCGCCACAGCCCTGCGCCGACACTGCGGCAAGACCTTCTCGAGAGCTGCACC 334

QY 71 tgatcgctcaccgcccgttgacacgagcgagccccaccagtgcccgcgagtgcgga 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 TGGTGACAGACCGGCGCATCCACAGGCGAGAGCCCTACGCTTGAGAGTGCGGCA 394

QY 131 agcgcttaccataaagccctatctgaacttcgacccgagcgatccacacgagcgagaagc 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 AGCGCTTCAGTGGAGCTGCACCTCATGACAGCACACCGCGCATCCACACGGGCGAGAGC 454

QY 191 cctaccggtgcaagagtgccgcccgcgcttcgagcacaacccaacctgctgtctaca 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 CTTACACTGCGCCGACTGCGGCGGCGAGCTTCACAGAGCAAGAGCTTGCCCAAGCAC 514

QY 251 gcaagattcaagaagatcgcgaggggtcgccagcgccccgcgcccggagagccccc 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 GCGCGTCGACAGGCGGCGCTCAAGCCCTTCGTGTGCCGCGCTGCGGCGCGGCTTCAGCC 574

QY 311 agctgcagccgcccagagagtcgcgagccgagcccaaccggcggtacctctgaac 370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 AGCCCAAGAGAGCTGCGGCTGACCTGCGGCTGACACCGGAGCTGTGCGGCGCTGCGCTGG 634

QY 371 cgagccagagagcgcccgccagagggcccgccagagcgaccccgagga 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 CGGCCAAGGCTGCTGCGGCTGACCTGCGGCGGCGCAAGGCGCCGA 680

```

Search completed: February 22, 2001, 03:16:44  
Job time: 15062 sec

**THIS PAGE BLANK (USPTO)**